H. pylori adhesin HopQ engages in a virulence-enhancing interaction with human CEACAMs

Anahita Javaheri[1,15,‡], Tobias Kruse[2,‡], Kristof Moonens[3,4,‡], Ayla Debraekeleer[3,4], Raquel Mejías-Luque[1,15], Isabell Asche[5], Nicole Tegtmeyer[5], Behnam Kalali[1,2], Nina C. Bach[6], Stephan A. Sieber[6], Darryl J. Hill[7], Verena Königer[8], Christof R. Hauck[9], Roman Moskalenko[10], Rainer Haas[8], Dirk H. Busch[1], Esther Klaile[11,12], Hortense Slevogt[11], Alexej Schmidt[13,14], Steffen Backert[5], Han Remaut[3,4,‡], Bernhard B. Singer[12‡] and Markus Gerhard[1,2,15‡*]

Affiliations:

[1]Institute for Medical Microbiology, Immunology and Hygiene; Technische Universität München; Munich, 81675, Germany,

[2]Imevax GmbH, 81675 Munich

[3]Structural and Molecular Microbiology, Structural Biology Research Center, VIB, Pleinlaan 2, 1050 Brussels, Belgium

[4]Structural Biology Brussels, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussels, Belgium

[5]Friedrich Alexander University Erlangen, Department of Biology, Division of Microbiology, Erlangen, Germany

[6]Center for Integrated Protein Science Munich, Department Chemie, Institute of Advanced Studies, Technische Universität München, 85747 Garching, Germany

[7]School of Cellular & Molecular Medicine, University of Bristol, BS8 ITD, Bristol, UK

[8]Max von Pettenkofer-Institut für Hygiene und Medizinische Mikrobiologie, Department of Bacteriology, Ludwig-Maximilians-Universität, D-80336 Munich, Germany

[9]Lehrstuhl für Zellbiologie, Universität Konstanz, Konstanz, Germany

[10]Department of Pathology, Sumy State University, Sumy 40000, Ukraine

[11]Septomics Research Centre, Jena University Hospital, 07745 Jena, Germany.

[12]Center for Sepsis Control and Care (CSCC), Jena University Hospital, 07747 Jena, Germany

[13]Institute of Anatomy, Medical Faculty, University Duisburg-Essen, 45122 Essen, Germany

[14]Department of Medical Biosciences, Pathology, Umeå University, SE-901 85 Umeå, Sweden

[15]German Center for Infection Research, Partner Site Munich, Munich, Germany

*Correspondence to: markus.gerhard@tum.de

‡ These authors contributed equally to this work

Summary: *Helicobacter pylori* specifically colonizes the human gastric epithelium and is the major causative agent for ulcer disease and gastric cancer development. Here we identified members of the carcinoembryonic antigen-related cell adhesion molecule (CEACAM) family as novel receptors of *H. pylori* and show that HopQ is the surface-exposed adhesin that specifically binds human CEACAM1, CEACAM3, CEACAM5 and CEACAM6. HopQ - CEACAM binding is glycan-independent and targeted to the N-domain. *H. pylori* binding induces CEACAM1 mediated signaling, and the HopQ-CEACAM1 interaction enables translocation of the virulence factor CagA into host cells, and enhances the release of pro-inflammatory mediators such as interleukin-8. Based on the crystal structure of HopQ, we found that a β-hairpin insertion (HopQ-ID) in HopQ's extracellular 3+4 helix bundle domain is important for CEACAM binding. A peptide derived from this domain competitively inhibits HopQ-mediated activation of the Cag virulence pathway, as genetic or antibody-mediated abrogation of the HopQ function shows. Together, our data imply the HopQ-CEACAM1 interaction as potentially promising novel therapeutic target to combat *H. pylori*-associated diseases.

Helicobacter pylori (*H. pylori*) is one of the most prevalent human pathogens, colonizing half of the world's population. Chronic inflammation elicited by this bacterium is the main cause of gastric cancer[1]. During co-evolution with it's human host over more than 60.000 years[2], the bacterium has acquired numerous adaptations for the long-term survival within its unique niche, the stomach. This includes the ability to buffer the extreme acidity of this environment, the interference with cellular signaling pathways, the evasion of the human immune response and a strong adhesive property to host cells[3]. Specifically, *H. pylori* persistence is facilitated by the binding of BabA and SabA adhesins to the human blood group antigen Leb and the sLex antigen, respectively[4-6]. However, adhesion to blood group antigens is not universal, is dynamically regulated during the course of infection and can also be turned off [7]. We observed that *H. pylori* was capable of binding to human gastric epithelium of non-secretors. Therefore, we hypothesized that the bacterium might be able to interact with other cell surface receptors to ensure persistent colonization.

We here show that the *H. pylori* adhesin HopQ specifically interacts with human carcinoembryonic antigen-related cell adhesion molecules (CEACAMs). CEACAMs embrace a group of immunoglobulin superfamily-related glycoproteins with a wide tissue distribution. CEACAM1 can be expressed in leukocytes, endothelial and epithelial cells, CEACAM3 and CEACAM8 in granulocytes, CEACAM5 and CEACAM7 in epithelial cells and CEACAM6 in epithelia and granulocytes. In epithelial cells, transmembrane anchored CEACAM1 as well as glycosylphosphatidylinositol-linked CEACAM5, CEACAM6 and CEACAM7 localize to the apical membrane[8]. CEACAMs modulate diverse cellular functions such as cell adhesion, differentiation, proliferation, and cell survival. Some CEACAMs were recognized as valuable tumor markers due to their enlarged expression in the malignant tissue and increased sera level[9]. In recent years, CEACAMs have also emerged as immunomodulatory mediators[10]. Interestingly, in humans, several CEACAMs have been found to specifically interact with bacteria such as *Neisseria, Haemophilus influenzae*, *Moraxella catarrhalis*, and *Escherichia coli* [11].

H. pylori binds to CEACAMs expressed in human stomach

Based on the observation that *H. pylori* efficiently colonizes individuals in the absence of Lewis blood group antigens[12] on the one hand, and the increased expression of members of the carcinoembryonic antigen-related cell adhesion molecule family (CEACAMs) in gastric tumors, we hypothesized that *H. pylori* may employ CEACAMs as receptors. Using pull down and flow cytometric approaches we found a robust interaction of the *H. pylori* strain

G27 with recombinant human CEACAM1-Fc (Fig. 1a), comparable to that of *Moraxella catarrhalis* (Extended Data Fig. 1a and b). As negative control, *Moraxella lacunata* did not bind to human CEACAM1, nor did *Campylobacter jejuni,* a pathogen closely related to *H. pylori* (Extended Data Fig. 1a and b). When testing for CEACAM specificity, we observed a clear interaction of *H. pylori* also with CEACAM3, 5 and 6, but not with CEACAM8 (Fig.1b and Extended Data Fig. 1c and d). Importantly, all *H. pylori* strains tested bound to these CEACAMs (Extended Data Fig.1f and g) including well-characterized reference strains (26695, J99) and the mouse-adapted strain SS1. However, binding strength differed among strains, with some preferentially binding to CEACAM1, and others to CEACAM5 and/or CEACAM6 (Extended Data Fig. 1f and g). We then analyzed the expression profiles of CEACAM1, CEACAM5 and CEACAM6 in normal and inflamed human stomach tissues and gastric cancer. If at all low levels of CEACAM1 and CEACAM5 were expressed at the apical side of epithelial cells, and their expression, as well as that of CEACAM6, was up-regulated upon gastritis and in gastric tumors (Fig. 1c and Extended Data Fig. 1e). During infection, *H. pylori*-induced responses may thus lead to increased expression of its CEACAM-receptors.

Adhesins from other bacteria were shown to specifically bind to the N-domain of human CEACAM1[13,14]. Similarly, we found that lack of the CEACAM1 N-domain abolished *H. pylori* binding completely (Fig. 1d). While for the interaction of *Neisseria meningitidis* with CEACAM1 the N-domain was necessary but not sufficient for binding[15], we observed binding of *H. pylori* to all tested CEACAM1 isoforms containing the N-domain, as well as to the N-domain alone (Fig. 1e). However, binding to the N-domain alone was weaker than to the N-A1-B CEACAM1 variant, which bound less than the N-A1-B-A2 variant (Fig. 1e and Extended Data Fig.1j), suggesting that these domains stabilize the CEACAM1-*H.pylori* interaction. Comparison of the respective N-domains indicated several residues conserved in CEACAM1, 5, and 6 but not in CEACAM8 (Extended Data Fig. 1h).

Species specificity of *Helicobacter* – CEACAM interaction

Although, murine and Mongolian gerbil models are routinely used to study gastric infection with *H. pylori*, the bacterium has been described so far to be naturally transmitted to only humans and non-human primates. Although CEACAMs are found in most mammalian species, and have a high degree of conservation, we found *H. pylori* to bind selectively to human, but not to mouse, bovine or canine CEACAM1 orthologues (Fig. 2a). However, we were surprised to find a strong interaction of *H. pylori* with rat-CEACAM1 (Fig. 2b and d). This interaction was also mediated through the N-domain of rat-CEACAM1 (Fig. 2c and d).

To substantiate these findings, we transfected human, mouse or rat-CEACAM1 into CHO cells, to which *H. pylori* does not adhere otherwise. Using confocal laser scanning microscopy, we observed *de novo* adhesion of *H. pylori* to CHO cells expressing human and rat, but not mouse CEACAM1 (Fig. 2e), which could be confirmed by pull down and Western blotting of lysates from transfected cells (Fig. 2f and Extended Data Fig. 2d). This finding makes *H. pylori* the first pathogen for which its CEACAM binding is not restricted to one species. Comparing the protein sequences of the CEACAM1-N domains, several amino acids conserved in human and rat differ in mouse (i.e. asn10, glu26, asn42, tyr48, pro59, thr66, asn77, val79, val89, ile90, glu103, tyr108) (Extended Data Fig. 2a). In addition, our findings of the lack of binding to mouse CEACAM1 may explain the differences seen in pathology between infected mice and humans[16].

The genus *Helicobacter* comprises several other spp. i.e. *H. felis, suis, and bizzozeronii* as well as the human pathogenic *H. bilis* and *H. heilmannii*. When assessing the interaction of these *Helicobacters* with human CEACAMs, only *H. bilis* bound to human CEACAM1, 5 and 6 (Extended Data Fig.2b and c). As *H. pylori*, *H. bilis* interacted with the N-domain of hu-CEACAM1 (Extended Data Fig.2b and c). This interaction may explain how *H. bilis* manages to colonize human bile ducts, where high levels of constitutively expressed CEACAM1 are present.

HopQ is the *Helicobacter* adhesin interacting with CEACAMs

In order to identify the CEACAM-binding partner in *Helicobacter,* we initially screened a number of *Helicobacter* mutants devoid of defined virulence factors that have been shown to be implicated in various modes of host cell interaction (BabA, SabA, AlpA/B, VacA, gGT, urease and the *cag*PAI)[5,6,17]. All of these mutants still bound to hu-CEACAM1 (Fig. 3a). Therefore we established an immunoprecipitation approach (Extended Data Fig. 3a) using *H. pylori* lysate and recombinant hu-CEACAM1-Fc coupled to protein G. Mass spectrometric analysis of the co-precipitate identified two highly conserved *H. pylori* outer membrane proteins as candidate CEACAM1 adhesins: HopQ and HopZ (Fig. 3b). Unlike a *hopZ* mutant, a *hopQ* deletion mutant was devoid of CEACAM1 binding (Fig. 3c). Importantly, the *hopQ* mutant was also unable to bind to CEACAM5 and 6 (Fig.3c).

Next we tested the binding of recombinant HopQ to different gastric cancer cell lines and found that HopQ interacted with AGS and MKN45 both endogenously expressing CEACAMs (Extended Data Fig.3b). HopQ did not bind to the CEACAM negative cell line MKN28. Utilizing our CHO transfectants, we found that the recombinant HopQ interacted

preferentially with CEACAM1 and 5, and to lesser extent to CEACAM3 and 6. No binding was observed to CHO cells expressing either CEACAM4, 7, or 8 (Extended Data Fig. 3c).

HopQ is a member of a *H. pylori*-specific family of outer membrane proteins, and shows no significant homology to other CEACAM-binding adhesins from other Gram-negative bacteria, i.e. Opa proteins or UspA1 from *Neisseria meningitidis* and *Neisseria gonorrhoeae* or *Moraxella catarrhalis*, respectively, and is therefore a novel bacterial factor hijacking CEACAMs. Like Opa and UspA1 [13,14], HopQ targets the N-terminal domain in CEACAMs, an interaction we found to require folded protein (see below) and was dependent on CEACAM sequence, resulting in specificity for human CEACAM1, 3, 5 and 6. The *H. pylori hopQ* gene (*omp27*; HP1177 in the *H. pylori* reference strain 26695) exhibits genetic diversity that represents two allelic families[18], type-I and type-II (Extended Data Fig. 3d), of which the type-I allele is found more frequently in *cag*(+)/s1-*vacA* type strains. Both alleles share 75 to 80% nucleotide sequences and exhibit a homology of 70% at the amino acid level[18]. Importantly, *hopQ* genotype shows a geographic variation, with the *hopQ* type-I alleles more prevalent in Asian compared to Western strains; and was also found to correlate with strain virulence, with type-I alleles associated with higher inflammation and gastric atrophy[19].

Structure and binding properties of the HopQ adhesin domain

HopQ belongs to a paralogous family of *H. pylori* outer membrane proteins (Hop's), to which also the blood group antigen binding adhesins BabA and SabA belong[5,6,17,20].To gain insight into its structure-function relationship we determined the binding properties and X-ray structure of a HopQ fragment corresponding to its predicted extracellular domain (residues 17-444 of the mature protein; $HopQ^{AD}$; Fig. 4a). $HopQ^{AD}$ showed strong, dose dependent binding to the N-terminal domain of human CEACAM1 (C1ND; residues 35-142) in ELISA (Fig. 4b) and isothermal titration calorimetry (ITC) revealed a 1:1 stoichiometry with a dissociation constant of 296±40 nM (Extended Data Fig. 4a). The $HopQ^{AD}$ X-ray structure shows that, like BabA and SabA, the HopQ ectodomain adopts a 3+4-helix bundle topology, though lacks the extended coiled-coil "stem" domain that connects the ectodomain to the transmembrane region (Fig. 4a and Extended Data Fig.4d). In BabA, the carbohydrate binding site resides fully in a 4-stranded β-domain that is inserted between helices 4 and 5[21] (Extended Data Fig.4d). In HopQ, a 2-stranded β-hairpin is found in this position (residues 180-218). Removal of the β-hairpin resulted in a soluble protein that showed a ~10 fold reduction of CEACAM1 binding affinity (Fig. 4b and Extended Data Fig. 4c), indicating that although the

HopQ insertion domain is implicated in binding, it does not comprise the full binding site as found in BabA (Fig. 4b).

The hitherto characterized Hop adhesins are lectins[5,6,17,22]. Instead, *H. pylori* was seen to retain binding to CEACAM1 upon enzymatic deglycosylation, and Far Western analysis revealed that $HopQ^{AD}$ specifically bound folded, but not denatured C1ND (Fig. 4c), suggesting HopQ-CEACAM binding relies on protein-protein rather than glycan-dependent interactions. Indeed, ITC binding profiles of $HopQ^{AD}$ titrated with non-glycosylated *E. coli* expressed C1ND (Ec-C1ND) revealed an equimolar interaction with a dissociation constant of 417±48 nM (Extended Data Fig. 4b), showing that CEACAM N-glycosylation only provides a minor stabilizing contribution to the HopQ-CEACAM interaction. To further map the HopQ binding site, we pre-incubated CEACAM1 with the *M. catarrhalis* adhesin UspA1, and found that this prevented binding by *H. pylori* (Fig. 4d), suggesting that both adhesins have overlapping binding epitopes. In further support, mutation of CEACAM1 residues Y34 or I91 within the UspA1 binding epitope reduced or nearly abrogated CEACAM1 binding by *H. pylori* (Fig. 4e). Interestingly, I91 is conserved in rat but mutated to T in mouse CEACAM1, possibly explaining the observed species specificity in HopQ binding (Extended Data Fig. 2a, see above).

HopQ – CEACAM1 interaction triggers cell responses

Available animal models only partially replicate the *H. pylori* pathogenesis observed in its human host and mouse CEACAMs did not support HopQ binding. Therefore, to further investigate how HopQ may influence adhesion and cellular responses, we sought to establish cellular pathogenesis models in which the HopQ-CEACAM mediated adhesion could be analyzed. According to Singer et al.[23], we characterized various gastric cell lines typically employed for *H. pylori in vitro* experiments regarding their expression of CEACAMs, and observed that MKN45, KatoIII and AGS did express CEACAM1, CEACAM5 and CEACAM6, whereas MKN28 showed no presence of CEACAMs (Extended Data Fig.5a and b). In parallel, CHO cells were stably transfected with CEACAM1-L (containing the immunoreceptor tyrosine-based inhibition motif (ITIM). Upon infection with *H. pylori* wild-type strain P12 and its isogenic *hopQ* deletion mutant, we observed a significantly reduced adherence to CHO-CEACAM1-L, MKN45 and AGS cells when hopQ was not present, while strains deficient in the adhesins BabA and SabA showed only slightly reduced adhesion (Fig. 5a and Extended Data Fig.5c). HopQ binding was also studied in human gastric biopsies from *H. pylori* infected individuals. Here, we detected that HopQ bound to the apical side human

gastric epithelium and co-localized with CEACAM in biopsies from *H. pylori* infected individuals (Fig. 5b and Extended Data Fig. 5d), while no binding was observed in CEACAM1 negative samples from normal stomach (not shown). In CHO-CEACAM1-L cells, we observed tyrosine-phosphorylation of the CEACAM1 ITIM domain upon exposure to *H. pylori*, which was apparent within 5 minutes, and was maintained for up to 1 hour (Fig.5c). Phosphorylation of the CEACAM1 ITIM domain is a well-known initial event triggering SHP1/2 recruitment inducing downstream signaling cascades[24,25]. Contact-dependent signaling through CEACAMs is a common means of modulating immune responses related to infection, inflammation and cancer[10], and these immune-dampening cascades likely reflect the multiple independent emergence of non-homologous CEACAM-interacting proteins in diverse mucosal Gram-negative pathogens including *Neisseria, Haemophilus, Escherichia, Salmonella, Moraxella* sp.[13,14]. For *H. pylori*, interaction with human CEACAM1 through HopQ may represent a critical parameter for immuno-modulatory signaling during colonization and chronic infection of man.

Additionally, *hopQ* mutant *H. pylori* strains showed an almost complete loss of *cag*PAI-dependent CagA translocation (Fig. 5d) and strongly reduced IL-8 induction (Fig.5e), while loss of other known adhesins had no effect on CagA delivery (Extended Data Fig.5e and f). This is in line with a previous study showing that in AGS gastric cancer cells, a *hop*Q mutant *H. pylori* strain exhibited reduced ability to activate NF-κB and altered translocation of CagA[26]. In contrast to our findings, Belogolova et al. did not observe reduced adherence of a *hopQ* mutant *H. pylori* P12 strain, which could be due to the observed growth dependent expression of CEACAMs in these cells.

To corroborate our data in an independent model and compensate for potential clonal effects in stably transfected cells, we transiently transfected HEK293 cells with human CEACAM (1-L,3,4,5,6,7,8) expression plasmids. Infection of these cells confirmed the defect in CagA translocation observed in CHO-CEACAM1-L cells, which was restored upon complementation of the *hopQ* mutant strain (P12Δ*hopQhopQ*$^{+}$) (Fig.5f and Extended Data Fig.5g). Also, cellular elongation, the so called "hummingbird phenotype", was significantly reduced upon deletion of *hopQ* (Fig. 5g and h). Further, we observed that *H. pylori* modulates important host transcription factors such as Myc or STAT3, in a *hopQ*-dependent fashion (Extended Data Fig. 5h). Our results reveal that HopQ-CEACAM binding leads to direct and indirect alterations in host cell signaling cascades, and start to shed light on these HopQ-associated virulence landscapes. Given the importance of these signaling events for gastric carcinogenesis, we explored if the CEACAM-HopQ interaction could be targeted in order to

prevent CagA translocation and downstream effects. Indeed, incubation of the cells with an α-CEACAM1 antibody, α-HopQ antiserum or a HopQ-derived peptide corresponding to the Hop-ID (aa 189-220) reduced CagA translocation in a dose dependent manner (Fig. 5i-k), but not corresponding controls (Extended Data Fig. 5h). These data demonstrate that the HopQ-CEACAM1 interaction is necessary for successful translocation of the oncoprotein CagA into epithelial cells as well as modulation of inflammatory signaling, and that interference with this interaction can prevent CagA translocation, giving an indication of the translational potential of HopQ targeting for *H. pylori* vaccination or immunotherapy.

Deletion of *hopQ* abrogates colonization in a rat model of *H. pylori* infection

As we have found binding of HopQ to human and rat, but not to mouse CEACAM, we finally determined the role of HopQ *in vivo,* using a rat model of *H. pylori* infection. Having observed that CEACAM1 was expressed in normal rat stomach (Fig. 6a and Extended Data Fig. 6b), we infected rats with the mouse adapted strain SS1, able to bind human and rat CEACAM1 (Extended Data Fig. 6a). While the wilt type SS1 was able to efficiently colonize rats, albeit at lower levels compared to the mouse, (Fig. 6b) , the *hopQ* deficient SS1 strain was not able to colonize rats at detectable levels, and could not induce an inflammatory response in comparison to the wild type SS1 strain (Fig. 6b and c). Therefore, in this model, HopQ seems also to serve as an important factor to mediate *H. pylori* colonization. While infection of rats with *H. pylori* has been described[27], our finding may allow the establishment of an animal model for studying *H. pylori* infection that better replicates the prevailing virulence pathways.

Discussion

The here identified CEACAM-binding property provides *H. pylori* a means of epithelial adherence in addition to the Lewis antigens used by the BabA and SabA adhesins[5,6,17]. While over-expression of CEACAMs in gastrointestinal tumors is well described, their up-regulation during *H. pylori*-induced inflammation in the stomach has not been reported so far, suggesting the pathogen has the ability to shape its own adhesive niche. A similar phenomenon has also been observed for the inflammation-induced up-regulation of sialylated antigens that form the receptors for the SabA adhesin[6]. A plausible route to CEACAM modulation is through the transcription factors NF-κB and AP1, both of which are induced during *H. pylori* infection[28] and are known to regulate CEACAM expression[29]. Though HopQ-dependent adherence may appear redundant to that of other adhesins like BabA, SabA or LabA, HopQ specializes on human CEACAMs and is required for *cag*PAI functionality. From the perspective of host-pathogen (i.e. human-*H. pylori*) co-evolution, the primary function of HopQ may lie in immune-modulation through CEACAM binding, and HopQ's indirect effects on other virulence cascades elicited by *H. pylori* such as that induced by increased CagA delivery may not have been initially "intended". The *cag*PAI was acquired by ancestral *H. pylori* in a single event that occurred before modern humans migrated out of East Africa around 58,000 years ago[30]. Thus, it is likely that the employment of CEACAM1 ligation by *H. pylori* occurred much earlier to support colonization and to modulate immune responses. This assumption is supported by the fact that all fully sequenced *H. pylori* strains bear *hop*Q (Extended Data Fig.3d), indicating that this is an essential outer membrane protein of *H. pylori*. Upon occurrence of type-I *H. pylori* strains by *cag*PAI acquisition more than 60,000 years ago[30] this ancient survival strategy was further implemented into a mechanism supporting pathogenicity, and thus may have contributed to the switch from commensal to pathogenic *H. pylori*[31]. Pathogenicity might even be further aggravated by our observation that CEACAMs are strongly up-regulated during gastritis, which further potentiates binding of *H. pylori* to epithelial cells and specifically facilitates CagA/cagPAI interaction with the host cells.

Taken together, the finding that *H. pylori* employs CEACAMs not only for bacterial adherence but also to induce cellular signaling may lead to a better understanding of the pathogenic mechanisms of these bacteria and might lead to novel therapeutic approaches to more effectively combat this highly prevalent infection and the associated gastric pathology.

Materials and Methods

Bacteria and bacterial growth conditions

The *H. pylori* strains G27[32], PMSS1[33], SS1[34], J99 (ATCC, 700824), 2808[35], 26695 (ATCC, 70039), TX30[36], 60190[37], P12[38], NCTC11637 (ATCC, 43504), Ka89 and *H. bilis* (ATCC43879) were grown on Wilkins–Chalgren blood agar plates under microaerobic conditions (10% CO2, 5% O2, 8.5% N2, and 37°C). *H. suis*[39] and *H. heilmannii*[40] were grown on Brucella agar and *H. felis* (ATCC 49179) and *H. bizzozeronii*[41] on brain-heart infusion (BHI) agar supplemented with 10% horse blood. *Moraxella catarrhalis* (ATCC, 25238) provided by C. R. Hauck (Konstanz Research School Chemical Biology, University of Konstanz, Germany), *Moraxella Lacunata* (ATCC 17967) and *Campylobacter jejuni* (ATCC, 33560) were cultured on brain–heart infusion (BHI) agar supplemented with 5% heated horse blood overnight at 37°C in a CO_2 incubator. The generation of an isogenic Δ*hopQ* mutant has been done by replacement of the entire gene by a chloramphenicol resistance cassette. For genetic complementation of *hopQ*, the 1,926 bp gene fragment of *H. pylori* strain P12 was amplified by PCR. This fragment was cloned into the complementation vector pSB1001 using the AphA3 cassette for selection. This fusion construct was introduced in the plasticity region of strain P12Δ*hopQ* (between ORFs HP0999 and HP1000) using a strategy as described[42].

Production of CEACAM proteins

The cDNA, which encodes the extracellular domains of human CEACAM1-Fc (consisting of N-A1-B1-A2 domains), human CEACAM1dN-Fc (consisting of A1-B1-A2, lacking the first 143 amino acids of the N-terminal IgV-like domain), rat CEACAM1-Fc (consisting of N-A1-B1-A2), rat CEACAM1dN-Fc (consisting of A1-B1-A2), human CEACAM3-Fc (consisting of N), human CEACAM6-Fc (consisting of N-A-B), human CEACAM8-Fc (consisting of N-A-B), respectively, were fused to a human heavy chain Fc-domain and cloned into the pcDNA3.1(+) expression vector (Invitrogen, San Diego, CA), sequenced and stably transfected into HEK293 (ATCC CRL-1573)cells as described[43]. The Fc chimeric CEACAM-Fc proteins were accumulated in serum-free Pro293s-CDM medium (Lonza) and were recovered by Protein A/G-Sepharose affinity Chromatography (Pierce). Proteins were analyzed by SDS-PAGE and stained by Coomassie blue demonstrating an equal amount and integrity of the produced fusion proteins (Extended Data Fig. 1i). Recombinant-human CEACAM5-Fc was ordered from Sino Biological Inc. The GFP-tagged CEACAMs (human-CEACAM1 and its variants, mouse-CEACAM1, bovine-CEACAM1 and canine-CEACAM1)

were provided by Dr. C. R. Hauck (University Konstanz, Germany). For production of the recombinant human CEACAM1 N-Domain (C1ND), the annotated domain (residues 35-142 of CEACAM1, Uniprot ID: P13688) was first backtranslated using the Gene Optimizer® (LifeTechnologies) and the leader sequence of the Igk-chain as well as a C-terminal Strep-Tag II was added. The gene was synthesized and seamlessly cloned into pCDNA3.4-TOPO (LifeTechnologies). Protein was produced in a 2 L culture of Expi293 cells according to the Expi293 expression system instructions (LifeTechnologies). The resulting supernatant was concentrated and diafiltered against ten volumes of 1x SAC buffer (100 mM Tris-HCl, 140 mM NaCl, 1 mM EDTA, pH 8.0) by crossflow-filtration, using a Hydrosart 5 kDa molecular-weight cutoff membrane (Sartorius). The retentate was loaded onto a StrepTrap HP column (GE Healthcare) and eluted with 1x SAC supplemented with 2.5 mM D-Desthiobiotin (IBA). The protein was stored at +4°C.

For the bacterial expression of the C1ND (Ec-C1ND) the amino acid sequence (residues 35-142 of CEACAM1, Uniprot ID: P13688) was codon optimized for expression in *E. coli*, synthesized by GeneArt *de novo* gene synthesis (Life Technologies), and cloned with a C-terminal His6 tag in the pDEST™14 vector using Gateway technology (Invitrogen). *E. coli* C43(DE3) cells were transformed with the resulting construct and grown in LB supplemented with 100 µg/mL ampicillin at 37°C while shaking. At OD_{600}=1 Ec-C1ND expression was induced with 1 mM IPTG overnight at 30°C. Cells were collected by centrifugation at 6.238 *g* for 15 minutes at 4°C and resuspended in 50mM Tris-HCl pH 7.4, 500 mM NaCl (4 mL/g wet cells) supplemented with 5 µM leupeptin and 1 mM AEBSF, 100 µg/mL lysozyme, and 20 µg/mL DNase I. Subsequently cells were lysed by a single passage in a Constant System Cell Cracker at 20 kPsi at 4 °C and debris was removed by centrifugation at 48.400 *g* for 40 minutes. The cytoplasmic extract was filtrated through a 0.45 µm pore filter and loaded on a 5 mL pre-packed Ni-NTA column (GE Healthcare) equilibrated with buffer A (50 mM Tris-HCl pH 7.4, 500 mM NaCl and 20 mM imidazole). The column was then washed with 40 bed volumes of buffer A and bound proteins were eluted with a linear gradient of 0-75 % buffer B (50 mM Tris-HCl pH 7.4, 500 mM NaCl and 500 mM imidazole). Fractions containing Ec-C1ND, as determined by SDS-PAGE, were pooled and concentrated in a 10 kDa MW cutoff spin concentrator to a final volume of 5 ml. To remove minor protein contaminants, the concentrated sample was injected onto the Hi-Prep™ 26/60 Sephacryl S-100 HR column (GE Healthcare) pre-equilibrated with a buffer containing 50 mM Tris-HCl pH 8.0, 150 mM NaCl. Fractions containing the Ec-C1ND complex were pooled and concentrated using a 10 kDa MW cutoff spin concentrator.

HopQ^{AD} and HopQ^{AD}ΔID cloning, production and purification

In order to obtain a soluble HopQ fragment, the HopQ gene from the *H. pylori* G27 strain (accession No. CP001173 Region: 1228696..1230621) HopQ fragment ranging from residues 37 – 463 was produced (residues 17-444 of the mature protein), thus removing the N-terminal β-strand and signal peptide, as well as the C-terminal β-domain expected to represent the TM domain. In HopQ^{AD}ΔID, the amino acids 184-212 of the mature protein were replaced by two glycines (Extended Data Fig.f). DNA coding sequences corresponding to the HopQ type I fragments was PCR-amplified from *H. pylori* G27 genomic DNA using primers (forward: GTTTAACTTTAAGAAGGAGATATACAAATGGCGGTTCAAAAAGTGAAAAACGC; reverse: TCAAGCTTATTAATGATGATGATGATGGTGGGCGCCGTTATTCGTGGTTG), containing 30bp overlap to the flanking target vector sequences of pPRkana-1, a derivative of pPR-IBA 1 (IBA GmbH) with the ampicillin resistance cassette replaced by the kanamycin resistance cassette, under a T7 promotor. In parallel, the vector was PCR-amplified using primers (forward: CACCATCATCATCATCATTAATAAGCTTGATCCGGCTGCTAAC ; reverse: GTTTAACTTTAAGAAGGAGATATACAAATG) as provided in table 1, using the same overlapping sequences in reversed orientation. The forward primer additionally carried the sequence for a 6x His-tag. The amplicons were seamlessly cloned using Gibson Assembly (New England Biolabs GmbH). Based on codon optimized HopQ^{AD} plasmid, the HopQ^{AD}ΔID constructs were cloned. The plasmids were amplified by 5' phosphorylated primers (forward: GGTGACGCTCAGAACCTGCTGAC; reverse: ACCACCTTTAGAGTTCAGCGGAG) replacing the ID region by two glycines, *Dpn*I (NEB) digested and blunt-end ligated by T4 ligase (NEB).

Escherichia coli BL21(DE3) cells (NEB GmbH) were transformed with the pPRkana-1 constructs, grown at 37°C with 275 rpm on auto-inducing terrific broth (TRB) according to Studier[44], supplemented with 2 mM $MgSO_4$, 100 mg/L Kanamycin-Sulfate (Carl Roth GmbH + Co. KG), 0.2 g/L PPG2000 (Sigma-Aldrich) and 0.2% w/v Lactose-monohydrate (Sigma-Aldrich), until an OD of 1-2 was reached. Afterwards, the temperature was lowered to 25°C and auto-induced overnight, reaching a final OD of 10-15 the following morning. Cells were harvested by centrifugation at 6000 *g* for 15 min at 4 °C using a SLA-3000 rotor in a Sorvall RC-6 Plus centrifuge (Thermo Fischer). Prior to cell disruption, cells were resuspended in 10 mL cold NiNTA buffer A (500 mM NaCl, 100 mM Tris-HCl, 25 mM Imidazole, pH 7.4) per gram of biological wet weight (BWW), supplemented with 0.1 mM AEBSF-HCl, 150 U/g

BWW DNase I and 5 mM $MgCl_2$and dispersed with an Ultra-Turrax T25 digital (IKA GmbH + Co. KG). Cell disruption was performed by high-pressure homogenization with a PANDA2000 (GEA NiroSoavi) at 800-1200 bar in 3 passages at 4 °C. The cell lysate was clarified by centrifugation at 25000 *g* for 30 min at 4 °C in a SLA-1500 rotor and remaining particles removed by filtration through a 0.2 µM filter.

HopQ fragments were purified by consecutive nickel affinity and size exclusion chromatography. Briefly, the clarified cell lysate was loaded onto a 5 mL pre-packed Ni-NTA HisTrap FF crude column (GE Healthcare) pre-equilibrated with buffer A, washed with ten column volumes (CV) of buffer A and the bound protein eluted with a 15 CV linear gradient to 75% NiNTA buffer B (500 mM NaCl, 100 mM Tris-HCl, 500 mM Imidazole, pH 7.4). Eluted peak fractions were collected, pooled and concentrated to a final concentration of 8-10 mg ml^{-1} using a 10 kDa molecular-weight cutoff spin concentrator. Subsequently, 5 mL of the concentrated protein were loaded onto a HiLoad 16/600 Superdex 75 pg column (GE Healthcare) pre-equilibrated with Buffer C (5 mM Tris-HCl, 140 mM NaCl, pH 7.3) and eluted at 1 mL/min. Finally, only protein corresponding to the monomer-peak was pooled and stored at +4 °C prior to crystallization. For analyzing the multimerization state of $HopQ^{AD}$, SEC was performed on a Superdex 200 10/300 GL (GE Healthcare) with 24 mL bed volume. The column was pre-equilibrated with Buffer C and subsequently, 25 µg protein injected and separated with a flow rate of 0.5 mL/min.

The HopQ interaction domain (HopQ-ID) representing peptide was HA-tagged, synthesized (EKLEAHVTTSKYQQDNQTKTTTSVIDTTNYPYDVPDYA) and HPLC purified (Peptide Specialty Laboratories, Heidelberg, Germany). For cellular assays, the lyophilized peptide was dissolved in sterile PBS to a concentration of 1 mM and dialysed with a 0.1-0.5 kDa molecular-weight cutoff membrane against PBS to remove remaining TFA. The peptide solution was stored at -20 °C until further use.

Detection of the HopQ-CEACAM interaction by ELISA

For detection of the interaction between CEACAM and $HopQ^{AD}$, recombinant C1ND (1 µg/mL) in PBS was coated over night at 4 °C onto a 96-well immunoplate (Nunc MaxiSorb). Wells were blocked with SmartBlock (Candor) for 2 h at RT. Subsequently, HopQ fragments were added in a fivefold series dilution ranging from 10 µg/mL to 0.05 ng/mL for 2h at room temperature. Next, anα-6xHis-HRP conjugate (clone 3D5, LifeTechnologies) was diluted 1:5000 and incubated for 1h at room temperature. For detection, 1-Step™ Ultra TMB-ELISA Substrate Solution (LifeTechnologies) was used and the enzymatic reaction was stopped with

2 N H_2SO_4. Washing (3-5x) in between incubation steps was carried out with PBS / 0.05% Tween20.

Isothermal titration calorimetry

ITC measurements were performed on a MicroCal iTC200 calorimeter (Malvern). 25 µM C1ND or EcC1ND were loaded into the cell of the calorimeter and 250 µM $HopQ^{AD}$ type I was loaded in the syringe. All measurements were done at 25°C, with a stirring speed of 600 rpm and performed in 20 mM HEPES buffer (pH 7.4), 150 mM NaCl, 5% (v/v) glycerol and 0.05% (v/v) Tween-20. Binding data were analyzed using the MicroCal LLC ITC200 software.

SDS-PAGE and native-PAGE for Western blot

CEACAM was separated with both SDS-PAGE and native-PAGE (resp. on 15% and 7.5% polyacrylamide gels) in ice-cold 25 mM Tris-HCl, 250 mM glycine buffer. Subsequently samples were transferred to PVDF-membranes by wet blotting at 25 V during 60 minutes in ice-cold transfer buffer (25 mM Tris-HCl, 250 mM glycine and 20% methanol). Membranes were blocked during one hour in 10% milk powder (MP), 1x PBS and 0.005% Tween-20. Both membranes were washed and incubated together in 5% MP, 1x PBS, 0.005% Tween-20 in presence of 2 µM $HopQ^{AD}$ type I for one hour to allow complex formation between $HopQ^{AD}$ I and CEACAM. After a washing step the C-terminal His-tag of HopQ (CEACAM is strep tagged) was detected by adding consecutively mouse α-His (AbDSerotec) and goat α-mouse antibody (Sigma-Aldrich) during respectively one hour and 30 minutes in 5% MP, 1x PBS, 0.005% Tween-20. After a washing step the blot was developed by adding BCIP/NBT substrate (5-bromo-4-chloro-3-indolyl-phosphate/nitro blue tetrazolium) (Roche) in developing buffer (10 mM Tris-HCl pH 9.5, 100 mM NaCl, 50 mM $MgCl_2$).

Bacterial pull down

Bacteria were grown overnight on WC dent agar plates. Bacteria were scraped from plates, suspended in PBS, and colony forming units (cfu) were estimated by optical density 600 readings according to a standard curve. Bacteria were washed twice with PBS and 2×10^8 cells/mL were incubated with soluble CEACAM-Fc or CEACAM-GFP proteins or CHO cell lysates for 1 h at 37 °C with head-over-head rotation. After incubation, bacteria were washed 5 times with PBS and either boiled in SDS sample buffer (62.5 mM Tris-HCl [pH 6.8], 2% w/v SDS, 10% glycerol, 50 mM DTT, and 0.01% w/v bromophenol blue) prior to SDS-PAGE and Western blotting or taken up in FACS buffer (PBS/0.5% BSA) for flow cytometry analysis.

Immunoprecipitation and Mass Spectrometry

Bacteria ($2x10^8$) in cold PBS containing protease and phosphatase inhibitors (Roche) were lysed by ultra-sonication on ice (10x, 20s). Cell debris was removed from the lysates by centrifugation at 15,000 rpm for 30 min at 4 °C, followed by pre-clearing with prewashed protein G-agarose (Roche Diagnostics). CEACAM1-Fc was added to the lysate (10 µg) and incubated for 1 h at 4 °C. Prewashed protein G-agarose (60 µL) were added to the antibody and lysate mixture and incubated 2 h at 4 °C. Beads were washed with PBS for five times to remove unspecifically bound proteins. Two-thirds of the beads were separated and used for mass spectrometry sample preparation. The supernatant was removed and the beads were resuspended twice in 50 µL 7M urea/ 2 M thiourea solved in 20 mM Hepes (pH 7.5) for denaturation of the proteins. Beads were pelleted by centrifugation and supernatants pooled and transferred to a new Eppendorf tube. Subsequently, proteins were reduced in 1 mM DTT for 45 min and alkylated at a final concentration of 5.5 mM iodacetamide for 30 min in the dark. The alkylation step was quenched by raising the DTT concentration to 5 mM for 30 min. All incubation steps were carried out at RT under vigorous shaking (Eppendorf shaker, 450 rpm). For digestion of the proteins 1 µL LysC (0.5 µg/µL) was added and the sample incubated for 4h at RT. To reduce the urea concentration the sample was diluted 1:4 with 50 mM triethylammonium bicarbonate and then incubated with 1.5 µL trypsin (0.5 µg/µL) at 37 °C over night. Trypsin was finally inactivated by acidification with formic acid. The supernatant was transferred to a new Eppendorf tube and pooled with the following wash fraction of the beads with 0.1% formic acid. The sample was adjusted to pH 3 with formic acid (100% v/v) and subjected to peptide desalting with a SepPak C18 column (50 mg, Waters). Briefly, the column was subsequently washed with 1 mL 100% acetonitrile and 500 µL 80% acetonitrile, 0.5% formic acid. The column was equilibrated with 1 mL 0.1% TFA, the sample was loaded and the column washed again with 1 mL 0.1% TFA. After an additional wash step with 500 µL 0.5% formic acid peptides were eluted twice with 250 µL 80% acetonitrile, 0.5% formic acid. The organic phase was then removed by vacuum centrifugation and peptides stored at -80 °C. Directly before measurement peptides were resolved in 20 µL 0.1% formic acid, sonified for 5 min (water bath) and the sample afterwards filtered with a prewashed and equilibrated filter (0.45 µm low protein binding filter, VWR International, LLC). Measurements were performed on an LC-MS system consisting of an Ultimate 3000 nano HPLC directly linked to an Orbitrap XL instrument (Thermo Scientific). Samples were loaded onto a trap column (2 µm, 100 A, 2 cm length) and separated on a 15 cm C18 column (2 µm, 100 A, Thermo Scientific) during a 150 min gradient ranging from 5

to 30% acetonitrile, 0.1% formic acid. Survey spectra were acquired in the orbitrap with a resolution of 60,000 at m/z 400. For protein identification up to five of the most intense ions of the full scan were sequentially isolated and fragmented by collision induced dissociation. The received data was analyzed with the Proteome Discoverer Software version 1.4 (Thermo Scientific) and searched against the *H. pylori* (strain G27) database (1501 proteins) in the SEQUEST algorithm. Protein N-terminal acetylation and oxidation of methionins were added as variable modifications, carbamidomethylation on cysteines as static modifications. Enzyme specificity was set to trypsin and mass tolerances of the precursor and fragment ions were set to 10 ppm and 0.8 Da, respectively. Only peptides that fulfilled X_{corr} values of 1.5, 2.0, 2.25 and 2.5 for charge states +1, +2, +3 and +4 respectively were considered for data analysis.

Cells, cell-bacteria co-culture and elongation phenotype quantitation assay

Gastric cancer cell lines MKN45[45], KatoIII (ATCC, HTB-103), MKN28 [46] and AGS (ATCC, CRL-1739) were obtained from ATCC and DSMZ, authenticated by utilizing Short Tandem Repeat (STR) profiling, cultured either sparse or to tight confluence in DMEM (GIBCO, Invitrogen, Carlsbad CA, USA) containing 2 mM L-glutamine (GIBCO, Invitrogen, CA, USA) supplemented with 10% FBS (GIBCO, Invitrogen, CA, USA) and 1% Penicillin/ Streptomycin (GIBCO, Invitrogen, CA, USA). All cell lines were maintained in an incubator at 37°C with 5% CO2 and 100% humidity, and were routinely mycoplasma-tested twice per year by DAPI stain and PCR. Plate-grown bacteria were suspended in DMEM and washed by centrifugation at 150 *g* for 5 min in a microcentrifuge. After resuspension in DMEM, the optical density at 600 nm was determined and bacteria were added to the overnight serum-deprived cells at different ratios of bacteria/cell (MOI) at 37°C to start the infection. After the indicated time, cells were washed twice with PBS and then lysed with 1% NP-40 in protease & phosphatase inhibitor PBS. HEK293 cells were chosen for CEACAM transfection studies because the cells were found to be negative for hu-CEACAM expression, and are easily transfectable. HEK cells were grown in 6-well plates containing RPMI 1640 medium (Invitrogen) supplemented with 25 mM HEPES buffer and 10% heat-inactivated FBS (Biochrom, Berlin, Germany) for 2 days to approximately 70% confluence. Cells were serum-deprived overnight and infected with *H. pylori* at MOI 50 for the indicated time points in each figure. After infection, the cells were harvested in ice-cold PBS containing 1 mM Na_3VO_4 (Sigma-Aldrich). Elongated AGS cells in each experiment were quantified in 5 different 0.25-mm^2 fields using an Olympus IX50 phase contrast microscope.

Transfection

A CHO cell line (ATCC) permanently expressing hu-CEACAM1-4L, mouse-CEACAM1-L and rat-CEACAM1-L were generated by stably transfecting cells with 4 µg pcDNA3.1-huCEACAM1-4L, pcDNA3.1-huCEACAM1-4S, pcDNA3.1-msCEACAM1-L, pcDNA3.1-ratCEACAM1-L plasmid (Singer), respectively, utilizing the lipofectamine 2000 procedure according to the manufacturer's protocol (Invitrogen). Stable transfected cells were selected in culture medium containing 1 mg/mL of Geniticinsulfat (G418, Biochrom, Berlin, Germany). The surface expression of CEACAM1 in individual clones growing in log phase was determined by flow cytometry (FACS calibur, BD). HEK293 cells were transfected with 4 µg of the HA-tagged CEACAM constructs or luciferase reporter constructs (Clontech, Germany) for 48 h with TurboFect reagent (Fermentas, Germany) according to the manufacturer's instructions.

Western blot

An equal volume of cell lysate was loaded on 8% SDS-PAGE gels and after electrophoresis, separated proteins were transferred to nitrocellulose membrane (Whatman/GE Healthcare, Freiburg, Germany). Membranes were blocked in 5% non-fat milk for 1 h at room temperature and incubated overnight with primary antibodies mAb 18/20 binding to CEACAM1,3,5, B3-17 and C5-1X (mono-specific for hu-CEACAM1, Singer), 4/3/17 (binding to CEACAM1,5, Genovac), and 5C8C4 (mono-specific for hu-CEACAM5, Singer), 1H7-4B (mono-specific for hu-CEACAM6, Singer), 6/40c (mono-specific for hu-CEACAM8, Singer), Be9.2 (α-rat-CEACAM1, kindly provided by Dr. W. Reutter, Charite, CBF, Germany), mAb 11-1H (α-rat-CEACAM1ΔN, Singer), phosphotyrosine antibody PY-99 (Santa Cruz, LaJolla, CA, USA), α-CagAphosphotyrosine antibody PY-972 [47], mouse monoclonal α-CagA antibody (Austral Biologicals, San Ramon, CA, USA), mouse monoclonal α-CEACAM1 (clone D14HD11Genovac/Aldevron, Freiburg,Germany) or goat α-GAPDH (Santa Cruz). After washing, membranes were incubated with the secondary antibody [HRP-conjugated α-mouse IgG (Promega)] and proteins were detected by ECL Western Blotting Detection reagents. The quantification was done by LabImage 1D software (INTAS).

Flow cytometry

The Fc-tagged CEACAMs (2.5 µg/mL) were incubated with *H. pylori* (OD_{600}=1) and subsequently with FITC-conjugated goat α-human IgG (Sigma-Aldrich). After washing with

FACS buffer, the samples were analyzed by gating on the bacteria (based on forward and sideward scatter) and measuring bacteria-associated fluorescence. In each case, 10,000 events per sample were obtained. Analysis was performed with the FACS CyAn (Beckman Coulter) and the data were evaluated with FlowJo software (Treestar). For the analysis of CEACAM mediated HopQ binding, indicated cell types (5×10^5 in 50 µL) were incubated with 20 µg/mL of *H. pylori* strain P12 derived,myc and 6x His-tagged recombinant HopQdiluted in 3% FCS/PBS for 1 h on ice. After three times washing with 3% FCS/PBS samples were labeled with 20 µg/mL of mouse α-c-mycmAb (clone 9E10, AbDSerotec) and subsequently with FITC conjugated goat α-mouse F(ab')2 (Dianova, Germany). In parallel, the presence of CEACAMs was controlled by staining cells utilizing the rabbit anti CEA pAb (A0115, Dianova) followed by FITC conjugated goat α-rabbit F(ab')2 (Dianova, Germany). Background fluorescence was determined using isotype-matched Ig mAb. The stained cell samples were examined in a FACScalibur flow cytometer (BD Biosciences, San Diego, CA) and the data were analyzed utilizing the CellQuest software. Dead cells, identified by PI staining, were excluded from the measurement.

Immunohistochemistry and Immunofluorescence

Following approval of the local ethics committee, paraffin-embedded human normal stomach, gastritis and cancer samples were randomly chosen from the tissue bank of the Institut für Pathologie, Klinikum Bayreuth Germany. Histological samples were excluded if tissue quality was poor. After antigen retrieval with 10 mM sodium citrate buffer pH 6 in pressure cooker, the sections were incubated with α-hu-CEACAM1, 5, 6 and α-rat-CEACAM1 antibodies (clone B3-17, 5C8C4, 1H7-4B and Be9.2, respectively). Sections were developed with SignalStain DAB (Cell Signaling) following manufacturer's instructions. Sections were counterstained with hematoxylin (Morphisto). The automated image acquisition was performed with Olympus Virtual Slide System VS120 (Olympus, Hamburg, Germany).

Visualization of the co-localization of HopQ and CEACAMs co-staining of normal and gastritis sections was performed utilizing HopQ-biotin followed by streptavidin-Cy3 and α-hu-CEACAM1, 3, 5, 6, 8 clone 6G5j followed by Alexa 488 coupled goat anti mouse antibody. The cell nuclei were stained with DAPI. DAPI and fluorescent proteins were analyzed with the Leica DMI4000B microscope.

Adherence assay

The adherence assay was performed according to Hytonen et al [48]. Briefly, human gastric epithelial cells (MKN45 and AGS) and CEACAM1-transfected CHO cells were grown in

antibiotic free DMEM (Gibco, Gaithersburg, MD) supplemented with 5% FCS and L-glutamine (2 mM, Sigma-Aldrich) on tissue culture 96 well plates (Bioscience) in 5% CO2 atmosphere for 2 days. To visualize *H. pylori* cells in adhesion assays, OD_{600}=1 of bacteria were fluorescence labeled with CFDA-SE (Molecular Probes) and washed with PBS. CFDA-SE was added at concentration of 10 μM for 30 min at 37˚C under constant rotation in the dark. Excess dye was removed by 3 times washing with PBS. Bacteria were resuspended in PBS until further use. Labelled bacteria were co-incubated (MOI 10) with the cells at 37°C with gentle agitation for 1 h. After washing with PBS (1 mL, ×3) to remove non-adherent bacteria, cells were fixed in paraformaldehyde (2%, 10 min). Bacterial binding was determined by measuring the percentage of cells that bound fluorescent-labeled bacteria using flow cytometry analysis.

IL-8 cytokine ELISA

AGS cell line was infected with *H. pylori* as described already and PBS-incubated control cells served as negative control. The culture supernatants were collected and stored at -20 °C until assayed. IL-8 concentration in the supernatant was determined by standard ELISA with commercially available assay kits (Becton Dickinson, Germany) according to described procedures.

HopQ-dependency of CagA virulence pathways

If not indicated otherwise, the AGS cell line (ATCC CRL-1730) was infected with the various *H. pylori* strains for 6 hours at a multiplicity of infection (MOI) of 50. The cells were then harvested in ice-cold PBS in the presence of 1 mM Na_3VO_4 (Sigma-Aldrich). In each experiment the number of elongated AGS cells was quantified in 10 different 0.25-mm^2 fields using a phase contrast microscope (Olympus IX50). CagA translocation was determined using the indicated antibodies detecting Tyr-phosphorylated CagA. All experiments were performed in triplicates. For inhibition experiments, cells were incubated with the indicated antibodies or peptides prior to infection.

Confocal microscopy

CHO cells were grown on chamber slides (Thermo Scientific), fixed in paraformaldehyde (4%, 10 min) and blocked with PBS/5% bovine serum albumin. CFDA-SE labelled bacteria (10 μM for 30 min at 37˚C under constant rotation in the dark) at MOI 5 were incubated with cells for 1 h at 37˚C under constant rotation. After 5X PBS washing, cell membranes were

stained with Deep Red (Life Technology) and cell nuclei with DAPI (Life Technology). Confocal images of cells were taken using a Leica SP5 confocal microscope.

Crystallization and structure determination of $HopQ^{AD}$

$HopQ^{AD}$was concentrated to 40 mg/mL and crystallized by sitting drop vapor diffusion at 20°C using 0.12 M alcohols (0.02 M 1,6-Hexanediol; 0.02 M 1-Butanol; 0.0 2M 1,2-Propanediol; 0.02 M 2-Propanol; 0.02 M 1,4-Butanediol; 0.02 M 1,3-Propanediol), 0.1 M Tris (base)/BICINE pH 8.5, 20% v/v PEG 500 MME; 10 % w/v PEG 20000 as a crystallization buffer. Crystals were loop-mounted and flash-cooled in liquid nitrogen. Data were collected at 100 K at beamline Proxima1 (SOLEIL, Gif-sur-Yvette, France) and were indexed, processed and scaled using the XDS package[49]. All crystals were in the $P2_1$ space group with approximate unit cell dimensions of a=57.7 Å, b=57.7 Å, c=285.7 Å and beta=90.1° and four copies of $HopQ_{442}$ per assymetric unit. Phases were obtained by molecular replacement using the BabA structure (PDB:5F7K)[21] and the program phaser[50,51]. The models were refined by iterative cycles of manual rebuilding in the graphics program COOT[52] and maximum likelihood refinement using Refmac5 [53]. Extended Data Table 2 summarizes the crystal parameters, data processing and structure refinement statistics.

Amino acid sequence alignment

The amino acid sequence alignment of the N-terminal domains of human, mouse and rat-CEACAM1 and human CEACAMs (1, 5, 6 and 8) was performed using CLC main Workbench (CLC bio).

Luciferase reporter assays

CHO-CEACAM1-L cells transfected with various luciferase reporter and control constructs (Clontech) were infected with *H. pylori* for 5 h and analyzed by luciferase assay using the Dual-Luciferase Reporter Assay System according to the manufactures instruction (Promega, USA). Briefly, cells were harvested by passive lysis, the protein concentration was measured with Precision Red (Cytoskeleton, USA) and the lysates were equalized by adding passive lysis buffer. The luciferase activity was measured by using a Plate Luminometer (MITHRAS LB940 from Berthold, Germany).

Animal experiments

Specific pathogen free, 120-150 g 4 weeks-old male Sprague Dawley rats, were obtained from Charles River Laboratories (Sulzfeld, Germany). Animals were randomly distributed into the different experimental groups by animal care takers not involved in the experiments, and

criteria for the exclusion of animals were pre-established. Investigator blinding was performed for all assessment of outcome and data, histology was performed by an independent investigator in a blinded manner. Animals were challenged twice intragastrically in groups of 8 with ~1 x 10^8 live *H. pylori* in 2 interval days. After 6 weeks infection, stomachs were removed and sectioned. One part was embedded in paraffin for histological analysis and another piece was weighted and homogenized to determine colony forming units (CFU)/mg stomach. Serial dilutions (1/10, 1/100 and 1/1000) were plated in WC dent plates. CFU were counted after 4 days.

The experiments were performed in the specific pathogen-free unit of Zentrum für Präklinische Forschung, Klinikum r. d. Isar der TU München, according to the allowance and guidelines of the ethical committee and state veterinary office (Regierung von Oberbayern, 55.2-1.54-2532-160-12).

Statistical Analysis

For in vitro experiments, normal distribution was determined by Shapiro–Wilk test. Normally distributed data were analyzed with two-tailed Student *t*-test or One-way ANOVA with post hoc Bonferroni test (comparing more than two groups) using Graph Pad Prism Software. Data are shown as mean ± s.e.m or S.D. for at least three independent experiments. P values <0.05 were considered significant. For animal studies, power calculation was performed based on previous animal experiments to achieve two sided significance of 0,05 while using lowest possible numbers to comply with the ethical guidelines for experimental animals. Mann-Whitney U test or ANOVA Kruskal-Wallis, Dunn's multiple comparison test were used to determine statistical significances.

References

1 Salama, N. R., Hartung, M. L. & Muller, A. Life in the human stomach: persistence strategies of the bacterial pathogen Helicobacter pylori. *Nature reviews. Microbiology* **11**, 385-399, doi:10.1038/nrmicro3016 (2013).

2 Atherton, J. C. & Blaser, M. J. Coadaptation of Helicobacter pylori and humans: ancient history, modern implications. *The Journal of clinical investigation* **119**, 2475-2487, doi:10.1172/JCI38605 (2009).

3 Montecucco, C. & Rappuoli, R. Living dangerously: how Helicobacter pylori survives in the human stomach. *Nature reviews. Molecular cell biology* **2**, 457-466, doi:10.1038/35073084 (2001).

4 Linden, S., Mahdavi, J., Hedenbro, J., Boren, T. & Carlstedt, I. Effects of pH on Helicobacter pylori binding to human gastric mucins: identification of binding to non-MUC5AC mucins. *The Biochemical journal* **384**, 263-270, doi:10.1042/BJ20040402 (2004).

5 Ilver, D. *et al.* Helicobacter pylori adhesin binding fucosylated histo-blood group antigens revealed by retagging. *Science* **279**, 373-377 (1998).

6 Mahdavi, J. *et al.* Helicobacter pylori SabA adhesin in persistent infection and chronic inflammation. *Science* **297**, 573-578, doi:10.1126/science.1069076 (2002).

7 Solnick, J. V., Hansen, L. M., Salama, N. R., Boonjakuakul, J. K. & Syvanen, M. Modification of Helicobacter pylori outer membrane protein expression during experimental infection of rhesus macaques. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 2106-2111, doi:10.1073/pnas.0308573100 (2004).

8 Hammarstrom, S. The carcinoembryonic antigen (CEA) family: structures, suggested functions and expression in normal and malignant tissues. *Seminars in cancer biology* **9**, 67-81, doi:10.1006/scbi.1998.0119 (1999).

9 Obrink, B. On the role of CEACAM1 in cancer. *Lung cancer* **60**, 309-312, doi:10.1016/j.lungcan.2008.03.020 (2008).

10 Gray-Owen, S. D. & Blumberg, R. S. CEACAM1: contact-dependent control of immunity. *Nature reviews. Immunology* **6**, 433-446, doi:10.1038/nri1864 (2006).

11 Voges, M., Bachmann, V., Kammerer, R., Gophna, U. & Hauck, C. R. CEACAM1 recognition by bacterial pathogens is species-specific. *BMC microbiology* **10**, 117, doi:10.1186/1471-2180-10-117 (2010).

12 Heneghan, M. A. *et al.* Effect of host Lewis and ABO blood group antigen expression on Helicobacter pylori colonisation density and the consequent inflammatory response. *FEMS immunology and medical microbiology* **20**, 257-266 (1998).

13 Virji, M., Watt, S. M., Barker, S., Makepeace, K. & Doyonnas, R. The N-domain of the human CD66a adhesion molecule is a target for Opa proteins of Neisseria meningitidis and Neisseria gonorrhoeae. *Molecular microbiology* **22**, 929-939 (1996).

14 Hill, D. J. & Virji, M. A novel cell-binding mechanism of Moraxella catarrhalis ubiquitous surface protein UspA: specific targeting of the N-domain of carcinoembryonic antigen-related cell adhesion molecules by UspA1. *Molecular microbiology* **48**, 117-129 (2003).

15 Kuespert, K., Roth, A. & Hauck, C. R. Neisseria meningitidis has two independent modes of recognizing its human receptor CEACAM1. *PloS one* **6**, e14609, doi:10.1371/journal.pone.0014609 (2011).

16 Peek, R. M. Helicobacter pylori infection and disease: from humans to animal models. *Disease models & mechanisms* **1**, 50-55, doi:10.1242/dmm.000364 (2008).

17 Icatlo, F. C., Goshima, H., Kimura, N. & Kodama, Y. Acid-dependent adherence of Helicobacter pylori urease to diverse polysaccharides. *Gastroenterology* **119**, 358-367 (2000).

18 Cao, P. & Cover, T. L. Two different families of hopQ alleles in Helicobacter pylori. *Journal of clinical microbiology* **40**, 4504-4511 (2002).

19 Ohno, T. *et al.* Relationship between Helicobacter pylori hopQ genotype and clinical outcome in Asian and Western populations. *J Gastroenterol Hepatol* **24**, 462-468, doi:10.1111/j.1440-1746.2008.05762.x (2009).

20 Alm, R. A. *et al.* Comparative genomics of Helicobacter pylori: analysis of the outer membrane protein families. *Infection and immunity* **68**, 4155-4168 (2000).

21 Moonens, K. *et al.* Structural Insights into Polymorphic ABO Glycan Binding by Helicobacter pylori. *Cell host & microbe* **19**, 55-66, doi:10.1016/j.chom.2015.12.004 (2016).

22 Rossez, Y. *et al.* The lacdiNAc-specific adhesin LabA mediates adhesion of Helicobacter pylori to human gastric mucosa. *The Journal of infectious diseases* **210**, 1286-1295, doi:10.1093/infdis/jiu239 (2014).

23 Singer, B. B. *et al.* Deregulation of the CEACAM expression pattern causes undifferentiated cell growth in human lung adenocarcinoma cells. *PloS one* **5**, e8747, doi:10.1371/journal.pone.0008747 (2010).

24 Muenzner, P., Bachmann, V., Zimmermann, W., Hentschel, J. & Hauck, C. R. Human-restricted bacterial pathogens block shedding of epithelial cells by stimulating integrin activation. *Science* **329**, 1197-1201, doi:10.1126/science.1190892 (2010).

25 Slevogt, H. *et al.* CEACAM1 inhibits Toll-like receptor 2-triggered antibacterial responses of human pulmonary epithelial cells. *Nature immunology* **9**, 1270-1278, doi:10.1038/ni.1661 (2008).

26 Belogolova, E. *et al.* Helicobacter pylori outer membrane protein HopQ identified as a novel T4SS-associated virulence factor. *Cell Microbiol* **15**, 1896-1912, doi:10.1111/cmi.12158 (2013).

27 Mahler, M. *et al.* Experimental Helicobacter pylori infection induces antral-predominant, chronic active gastritis in hispid cotton rats (Sigmodon hispidus). *Helicobacter* **10**, 332-344, doi:10.1111/j.1523-5378.2005.00320.x (2005).

28 Chang, Y. J. *et al.* Mechanisms for Helicobacter pylori CagA-induced cyclin D1 expression that affect cell cycle. *Cell Microbiol* **8**, 1740-1752, doi:10.1111/j.1462-5822.2006.00743.x (2006).

29 Muenzner, P., Naumann, M., Meyer, T. F. & Gray-Owen, S. D. Pathogenic Neisseria trigger expression of their carcinoembryonic antigen-related cellular adhesion molecule 1 (CEACAM1; previously CD66a) receptor on primary endothelial cells by activating the immediate early response transcription factor, nuclear factor-kappaB. *The Journal of biological chemistry* **276**, 24331-24340, doi:10.1074/jbc.M006883200 (2001).

30 Olbermann, P. *et al.* A global overview of the genetic and functional diversity in the Helicobacter pylori cag pathogenicity island. *PLoS genetics* **6**, e1001069, doi:10.1371/journal.pgen.1001069 (2010).

31 Suerbaum, S. & Josenhans, C. Helicobacter pylori evolution and phenotypic diversification in a changing host. *Nature reviews. Microbiology* **5**, 441-452, doi:10.1038/nrmicro1658 (2007).

32 Baltrus, D. A. *et al.* The complete genome sequence of Helicobacter pylori strain G27. *Journal of bacteriology* **191**, 447-448, doi:10.1128/JB.01416-08 (2009).

33 Arnold, I. C. *et al.* Tolerance rather than immunity protects from Helicobacter pylori-induced gastric preneoplasia. *Gastroenterology* **140**, 199-209, doi:10.1053/j.gastro.2010.06.047 (2011).

34 Lee, A. *et al.* A standardized mouse model of Helicobacter pylori infection: introducing the Sydney strain. *Gastroenterology* **112**, 1386-1397 (1997).

35 Lundin, A. *et al.* The NudA protein in the gastric pathogen Helicobacter pylori is an ubiquitous and constitutively expressed dinucleoside polyphosphate hydrolase. *J Biol Chem* **278**, 12574-12578, doi:10.1074/jbc.M212542200 (2003).

36 Atherton, J. C. *et al.* Mosaicism in vacuolating cytotoxin alleles of Helicobacter pylori. Association of specific vacA types with cytotoxin production and peptic ulceration. *The Journal of biological chemistry* **270**, 17771-17777 (1995).

37 Cover, T. L., Dooley, C. P. & Blaser, M. J. Characterization of and human serologic response to proteins in Helicobacter pylori broth culture supernatants with vacuolizing cytotoxin activity. *Infect Immun* **58**, 603-610 (1990).

38 Backert, S., Muller, E. C., Jungblut, P. R. & Meyer, T. F. Tyrosine phosphorylation patterns and size modification of the Helicobacter pylori CagA protein after translocation into gastric epithelial cells. *Proteomics* **1**, 608-617, doi:10.1002/1615-9861(200104)1:4<608::AID-PROT608>3.0.CO;2-G (2001).

39 Vermoote, M. *et al.* Genome sequence of Helicobacter suis supports its role in gastric pathology. *Vet Res* **42**, 51, doi:10.1186/1297-9716-42-51 (2011).

40 Haesebrouck, F. *et al.* Non-Helicobacter pylori Helicobacter species in the human gastric mucosa: a proposal to introduce the terms H. heilmannii sensu lato and sensu stricto. *Helicobacter* **16**, 339-340, doi:10.1111/j.1523-5378.2011.00849.x (2011).

41 Schott, T., Kondadi, P. K., Hanninen, M. L. & Rossi, M. Comparative genomics of Helicobacter pylori and the human-derived Helicobacter bizzozeronii CIII-1 strain reveal the molecular basis of the zoonotic nature of non-pylori gastric Helicobacter infections in humans. *BMC Genomics* **12**, 534, doi:10.1186/1471-2164-12-534 (2011).

42 Tegtmeyer, N. *et al.* Characterisation of worldwide Helicobacter pylori strains reveals genetic conservation and essentiality of serine protease HtrA. *Molecular microbiology* **99**, 925-944, doi:10.1111/mmi.13276 (2016).

43 Singer, B. B. *et al.* Soluble CEACAM8 interacts with CEACAM1 inhibiting TLR2-triggered immune responses. *PLoS One* **9**, e94106, doi:10.1371/journal.pone.0094106 (2014).

44 Studier, F. W. Protein production by auto-induction in high density shaking cultures. *Protein expression and purification* **41**, 207-234 (2005).

45 Hojo, H. & Onishi, Y. [Case suspected to be atypical diffuse myeloma]. *Nihon rinsho. Japanese journal of clinical medicine* **35**, 2659-2662 (1977).

46 Romano, M., Razandi, M., Sekhon, S., Krause, W. J. & Ivey, K. J. Human cell line for study of damage to gastric epithelial cells in vitro. *The Journal of laboratory and clinical medicine* **111**, 430-440 (1988).
47 Mueller, D. *et al.* c-Src and c-Abl kinases control hierarchic phosphorylation and function of the CagA effector protein in Western and East Asian Helicobacter pylori strains. *The Journal of clinical investigation* **122**, 1553-1566, doi:10.1172/JCI61143 (2012).
48 Hytonen, J., Haataja, S. & Finne, J. Use of flow cytometry for the adhesion analysis of Streptococcus pyogenes mutant strains to epithelial cells: investigation of the possible role of surface pullulanase and cysteine protease, and the transcriptional regulator Rgg. *BMC Microbiol* **6**, 18, doi:10.1186/1471-2180-6-18 (2006).
49 Krauth-Siegel, R. L. *et al.* Crystallization and preliminary crystallographic analysis of trypanothione reductase from Trypanosoma cruzi, the causative agent of Chagas' disease. *FEBS letters* **317**, 105-108 (1993).
50 Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta crystallographica. Section D, Biological crystallography* **67**, 235-242, doi:10.1107/S0907444910045749 (2011).
51 McCoy, A. J. *et al.* Phaser crystallographic software. *Journal of applied crystallography* **40**, 658-674, doi:10.1107/S0021889807021206 (2007).
52 Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta crystallographica. Section D, Biological crystallography* **66**, 486-501, doi:10.1107/S0907444910007493 (2010).
53 Murshudov, G. N. *et al.* REFMAC5 for the refinement of macromolecular crystal structures. *Acta crystallographica. Section D, Biological crystallography* **67**, 355-367, doi:10.1107/S0907444911001314 (2011).

Acknowledgments

We thank Jeannette Koch, Judith Lind, Birgit Maranca-Hüwel and Bärbel Gobs-Hevelke for their excellent technical support; Carolin Konrad, Johannes Fischer for support with rat experiments and Marie Roskrow for fruitful discussion and revision. KM and HR acknowledge use of the Soleil synchrotron, Gif-sur-Yvette, France under proposal 20131370 and support by VIB and the Flanders Science Foundation (FWO) through the Odysseus program, a postdoctoral fellowship and Hercules funds UABR/09/005. This work was supported by the German Centre for Infection Research, partner site Munich, to MG, by the BMBF 01EO1002 to E.K., the Mercator Research Center Ruhr An2012-0070 to BBS, the German Science Foundation CRC-796 (B10) and CRC-1181 (A04) to SB, the Collaborative Research Center/Transregio 124, Project A5 to HS.

Author Contribution

A.J., T.K., K.M., N.T., B.K., N.B., A.S. and B.B.S performed the experiments, B.B.S, R.H., V.K., E.K., H.S. and C.R.H. provided reagents and tools, A.J., B.B.S, H.R., D.B., R.M.-L., S.B. and M.G. conceived the experiments, analyzed the data and wrote the manuscript. All authors read and approved the final manuscript.

Author information

Reprints and permissions information is available at www.nature.com/reprints. M.G., B.K. and T.K. are employees and Shareholders of Imevax GmbH. M.G., A.J., B.S., S.B. and T.K. are named as inventors on a patent application regarding HopQ. The other authors declare no conflict of interest. Correspondence and requests for materials should be addressed to markus.gerhard@tum.de.

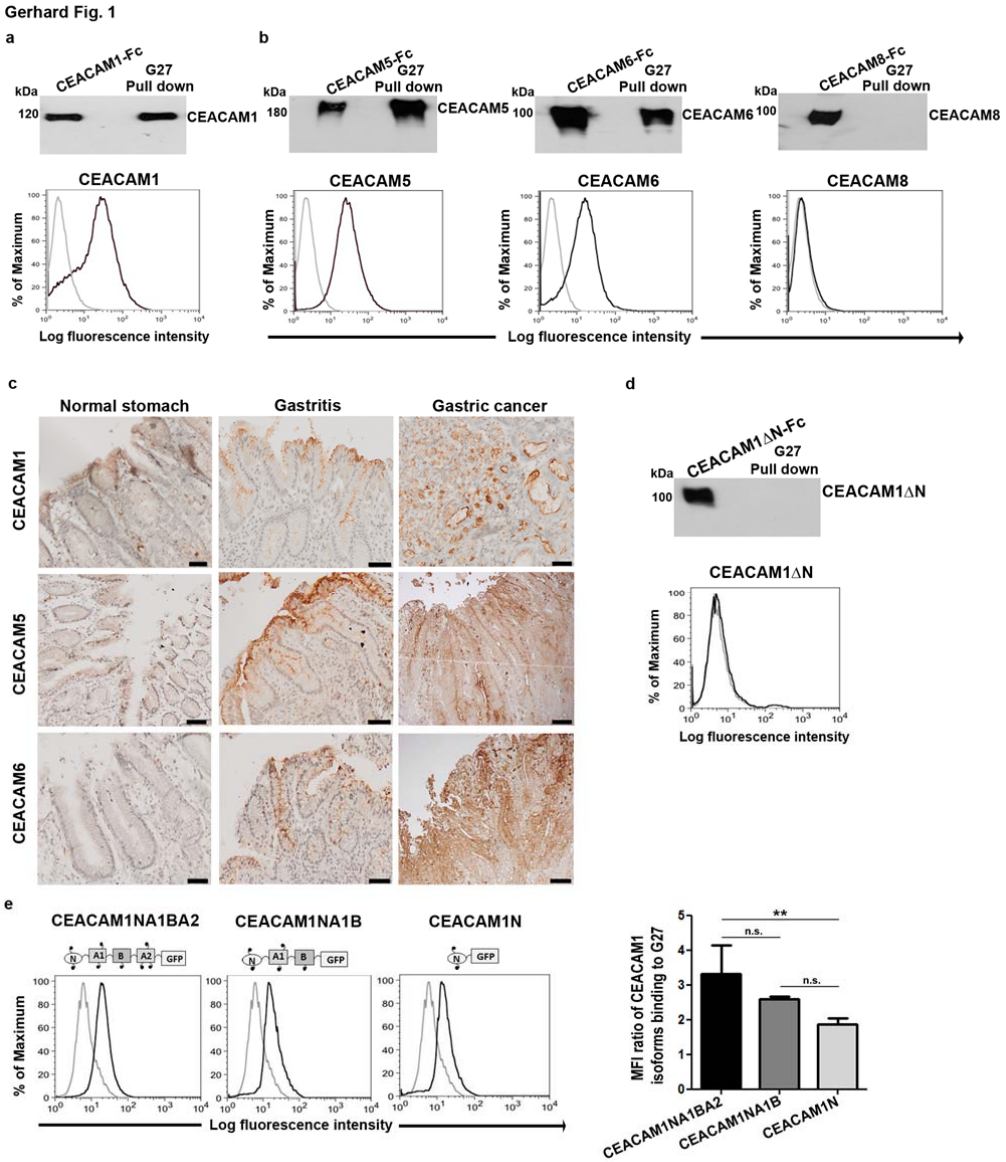
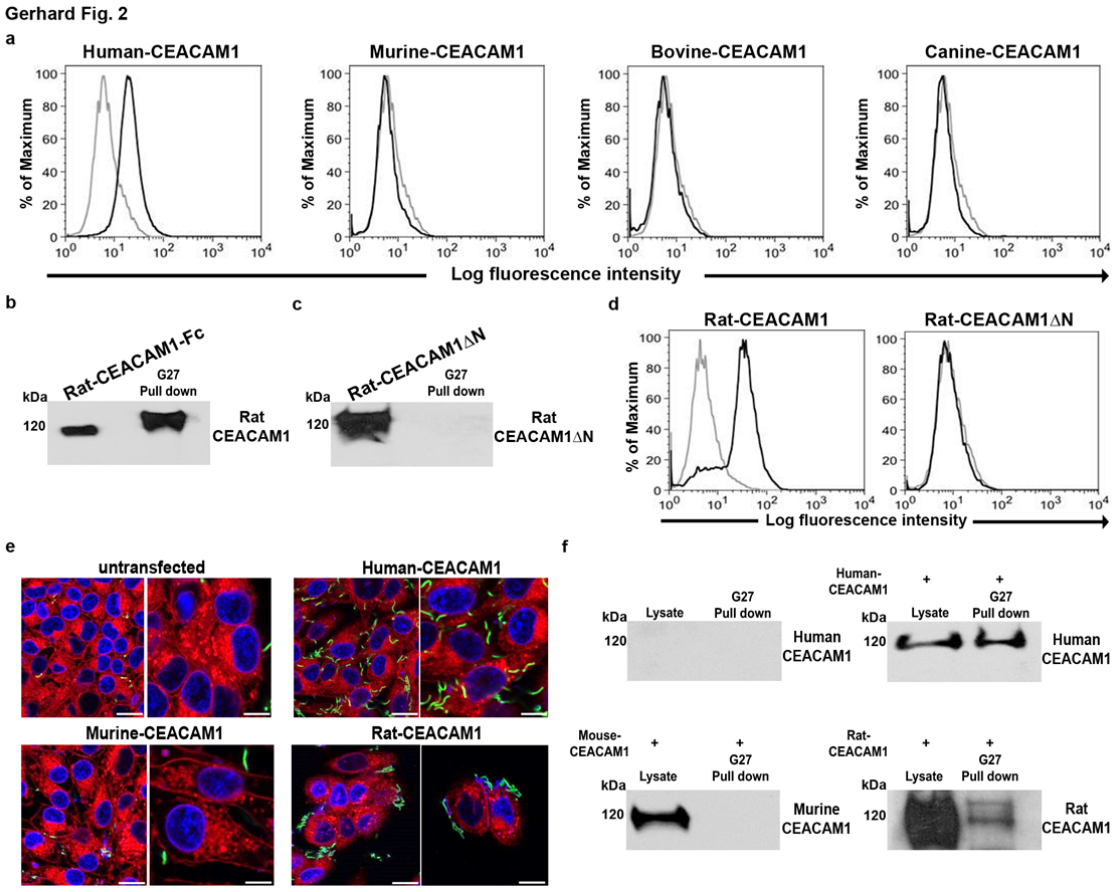


Figure 1 *H. pylori* employs the N-terminal domain of hu-CEACAM1 and binds CEACAM5 and CEACAM6 but not CEACAM8. *H. pylori* G27 strain binding to human CEACAM1-Fc (a) and human CEACAM5-Fc, CEACAM6-Fc or CEACAM8-Fc (b) was analyzed by pull down experiments followed by western blot analysis and flow cytometry (n=3). (c) CEACAM1, CEACAM5 and CEACAM6 expression detected by immunohistochemistry in human normal stomach, gastritis and gastric cancer samples**.** Scale bars, 50 µm. (d) Binding of *H. pylori* to human CEACAM1ΔN-Fc (lacking the complete N-domain) detected by western blot after pull down or by flow cytometry. One representative experiment of 4 is shown. (e) *H. pylori* binding to CEACAM variants analyzed by flow cytometry. Mean Fluorescence Intensity (MFI) ratios (mean, S.D.) are shown (n=4). One-way ANOVA, *P* value= 0.009, n. s.: not significant.

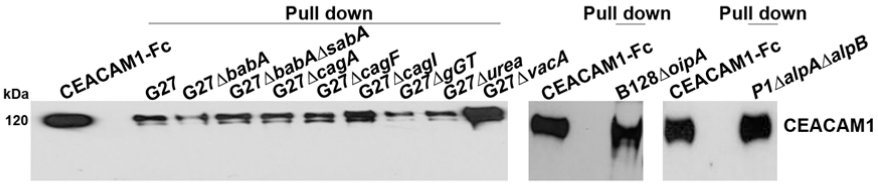


878
879

880 **Figure 2 *H. pylori* binding to CEACAM1 orthologues.** (a) *H. pylori* G27 strain binding to
881 human, murine, bovine and canine CEACAM1 determined by flow cytometry. (b) and (c) *H.*
882 *pylori* (G27) binding to rat-CEACAM1-Fc (b) and rat-CEACAM1ΔN-Fc (c) detected by
883 western blot after bacterial pull down. (d) Binding of G27 *H. pylori* strain to rat-CEACAM1
884 and rat-CEACAM1ΔN detected by flow cytometry. (e) Representative confocal images of *H.*
885 *pylori* binding to human, rat and mouse CEACAM1-expressing CHO cells. Untransfected
886 CHO served as control. Scale bars: left panels, 25 µm, right panels, 10 µm. (f) *H. pylori* G27
887 pull down of whole cell lysates of untransfected, human-, mouse- and rat CEACAM1-
888 transfected CHO cells. CEACAM1 was detected using species-specific CEACAM1
889 antibodies, as indicated. Representative experiments are shown (n=3).

Gerhard Fig. 3

a

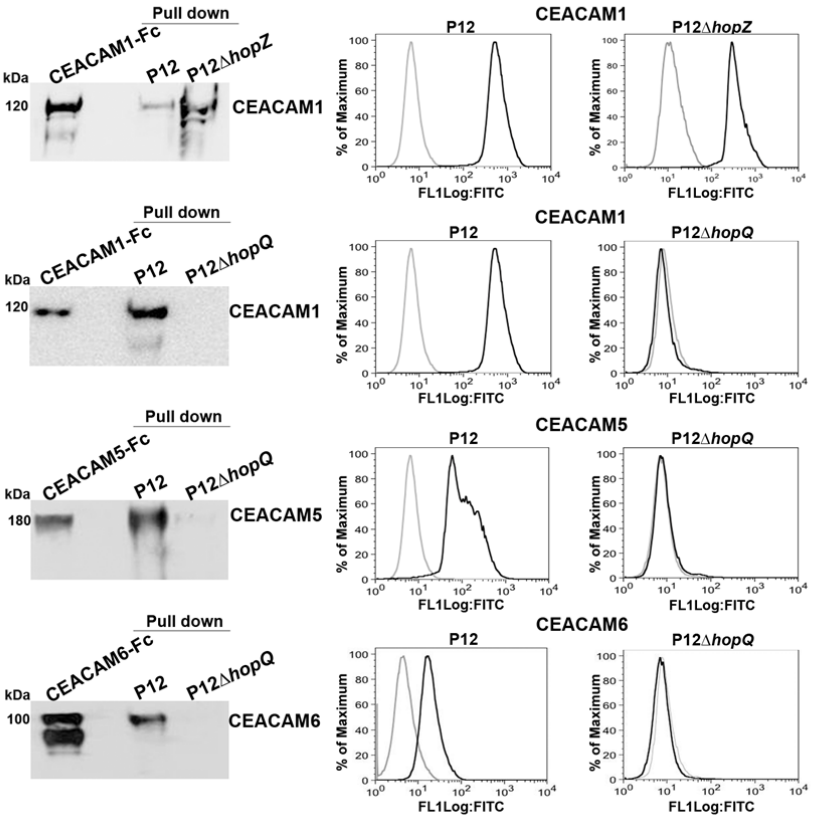


b

Mass spectrometry analysis

Accession	Description	Score	Unique Peptides	PSMs	Area	AAs	MW [kDa]
B5Z8H1	Outer membrane protein hopQ	34.29	5	9	4.714E8	641	69.9
B5Z6D0	Outer membrane protein hopZ	31.74	4	8	7.565E8	666	72.4
B5Z783	Outer membrane protein sabB	23.37	5	7	8.180E8	638	70.7
B5Z7R8	Outer membrane protein hopC	14.40	3	3	6.037E8	517	56.2
B5ZA84	Outer membrane protein babA	8.0	1	2	4.238E8	744	80.8
B5Z786	Outer membrane protein sabA	6.66	2	2	2.580E8	651	72.1

c



890

891 **Figure 3 *H. pylori* binds to CEACAM1 via HopQ.** (a) Human CEACAM1 detected by
892 western blot after pull down of various *H. pylori* G27 knockout strains incubated with human
893 CEACAM1-Fc. (b) Candidate outer membrane proteins of *H. pylori* strain G27 binding to
894 human CEACAM1-Fc (for complete MS table see Suppl. Table 1). (c) *H. pylori* strains P12,
895 P12Δ*hopQ* and P12Δ*hopZ* binding to hu-CEACAM1-, CEACAM5- and CEACAM6-Fc
896 detected by western blot and FACS analysis after pull down. Representative experiments are
897 shown (n=3).

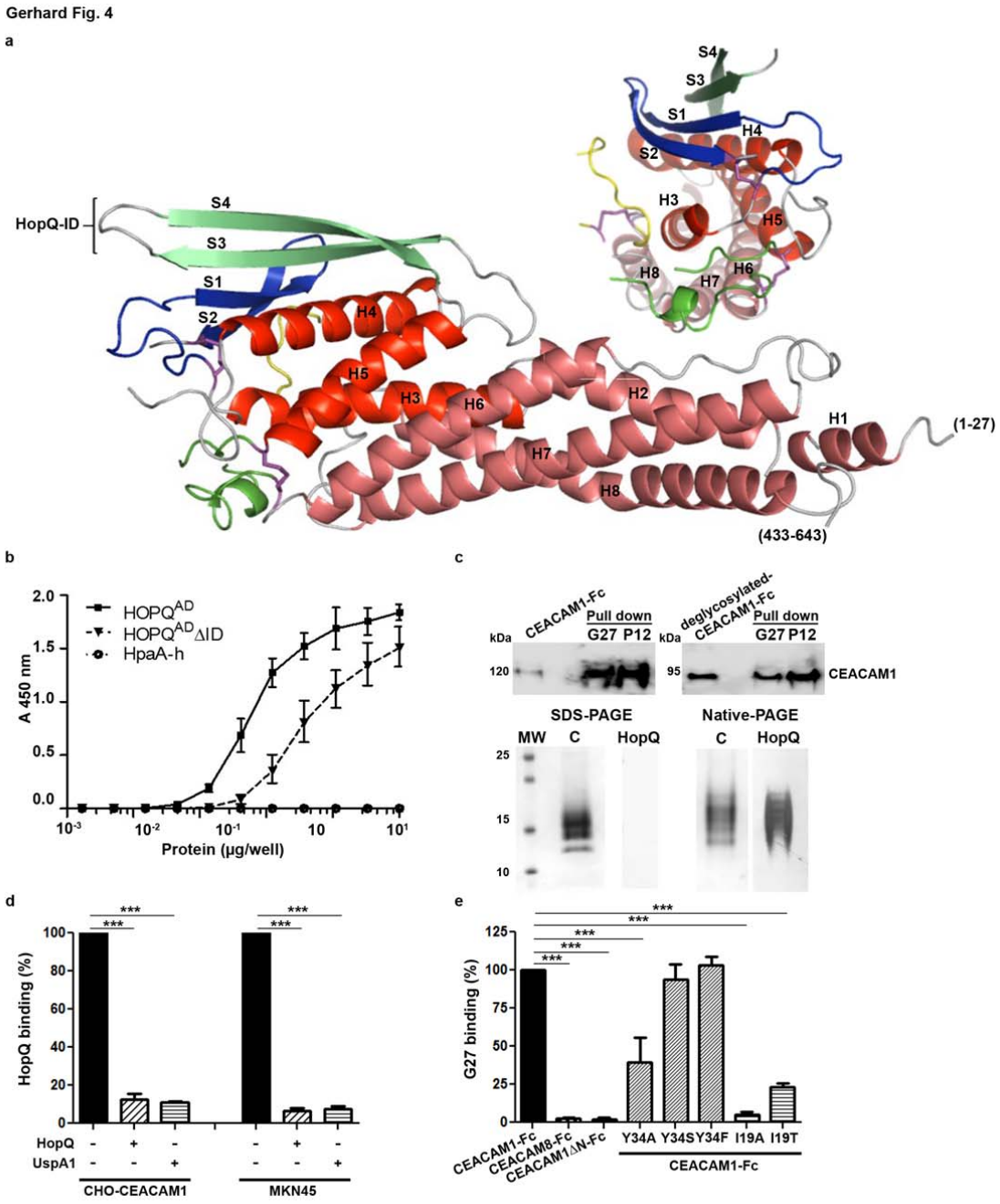


Figure 4. X-ray structure and binding properties of the HopQ adhesin domain. (a) Ribbon representation of the HopQ^{AD} showing the 3+4-helix bundle topology (colored red and brick, respectively). Three Cys pairs (Cys102-Cys131, Cys237-Cys269 and Cys361-Cys384) conserved in most Hop family members pinch off extended loops are colored blue, yellow and green. HopQ-ID; green, β-hairpin insertion. (b) ELISA titers of HopQ^{AD} or mutant HopQ^{AD} lacking the HopQ-ID (HopQ^{AD}ΔID) binding to increasing concentrations of C1-N domain (C1ND) (*n*=4, mean, S.D.). (c) Upper panel, pull down experiments of *H. pylori* strains incubated with de-glycosylated human CEACAM1-Fc. Lower panel, SDS and native

907 page of C1ND stained with Coomassie-blue ("C") or with $HopQ^{AD}$ in a far western blot
908 ("HopQ") experiment. (d) HopQ binding (%) to CEACAM1 in CHO and MKN45 cells after
909 pre-incubation with recombinant HopQ or UspA1, respectively. Mean, S.D. of three
910 independent experiments are shown. (e) *H. pylori* G27 binding (%) to CEACAM1,
911 CEACAM1ΔN and different CEACAM1 variants. CEACAM8 was used as negative control.
912 Mean, S.D. of three independent experiments are shown. One-way ANOVA with Bonferroni's
913 correction for multiple comparisons. ***$P \leq 0.001$.
914
915
916
917
918

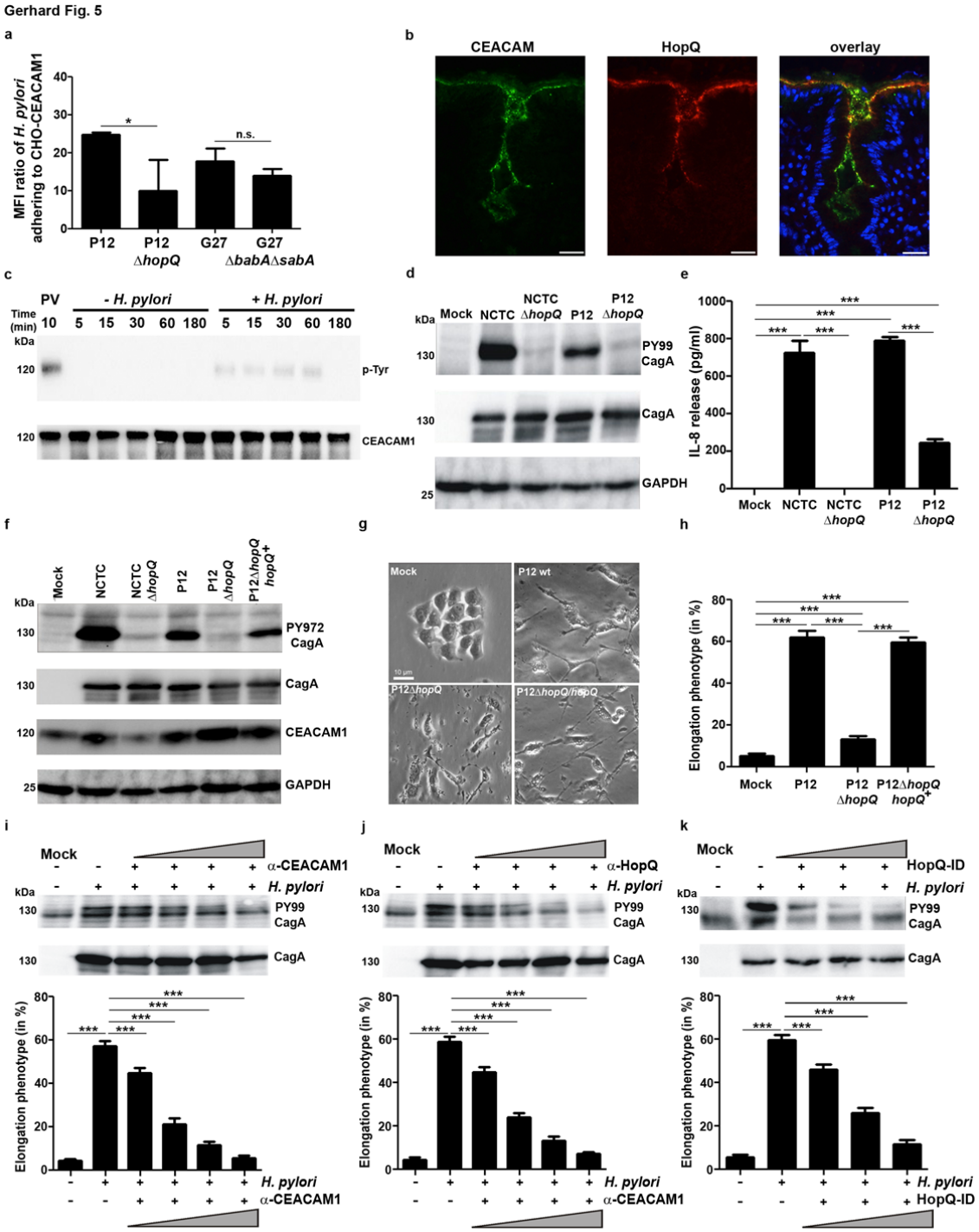


Figure 5 Deletion of *hopQ* in *H. pylori* leads to reduced bacterial cell adhesion and abrogates CagA delivery, IL-8 release and cell elongation. (a) *H. pylori* binding to CHO-hu-CEACAM1-L cells detected by flow cytometry analysis (n=3). Means ± S.D. are shown. Two-tailed *t*-test, * $P \leq 0.03$. (b) Immunofluorescence detection of apical CEACAM expression (green) and HopQ binding (red) in the gastric epithelium from human gastritis biopsies. Scale bar 25 µm. (c) CEACAM1 Tyr-phosphorylation and total CEACAM1 levels in

uninfected and *H. pylori*-infected CHO-CEACAM1-L cells. Pervanadate (PV) treatment served as positive control. (d) CagA phosphorylation detected in lysates of AGS cells after infection with *H. pylori* P12, NCTC11637 and corresponding isogenic *hopQ* mutants (e) Secreted IL-8 by AGS cells after infection with the indicated *H. pylori* strains (mean, S.D. of three independent experiments are shown). One-way ANOVA with Bonferroni's correction for multiple comparisons. ***P≤0.001. (f) CagA phosphorylation and CEACM1 levels in HA-tagged HEK293-hu-CEACAM1 transfectants infected with indicated *H. pylori* strains. (g) Representative phase contrast micrographs of AGS cells infected for 6 h with P12, P12Δ*hopQ* or P12Δ*hopQhopQ*$^{+}$ re-expressing wt *hopQ* gene. (h) Quantification of elongation phenotype induced in AGS cells after infection with the indicated *H. pylori* strains. Data (mean, S.D.) of three independent experiments are shown. One-way ANOVA with Bonferroni's correction for multiple comparisons. ***P≤0.001. (i) CagA phosphorylation and quantification of the elongation phenotype (five different 0.25-mm^2 fields) after *H. pylori* P12 infection of AGS cells pre-treated with 2, 5, 10 or 20 µg of α-CEACAM Ab (lanes 3-6). Data (mean, S.D.) of three independent experiments are shown. One-way ANOVA with Bonferroni's correction for multiple comparisons. ***P≤0.001. (j) CagA phosphorylation and quantification of the elongation phenotype after infection of AGS with wild type *H. pylori* pre-treated with 2, 5, 10 or 20 µg of α-HopQ (lanes 3-6) Data (mean, S.D.) of three independent experiments are shown. One-way ANOVA with Bonferroni's correction for multiple comparisons. ***P≤0.001. (k) CagA phosphorylation in *H. pylori*-infected AGS cells pre-incubated with a HopQ-derived peptide (1 µM, 2.5 µM and 5 µM) corresponding to the HopQ-ID (aa 189-220). Cell elongation (mean, S.D.) from 3 independent experiments is shown. One-way ANOVA with Bonferroni's correction for multiple comparisons. ***P≤0.001.

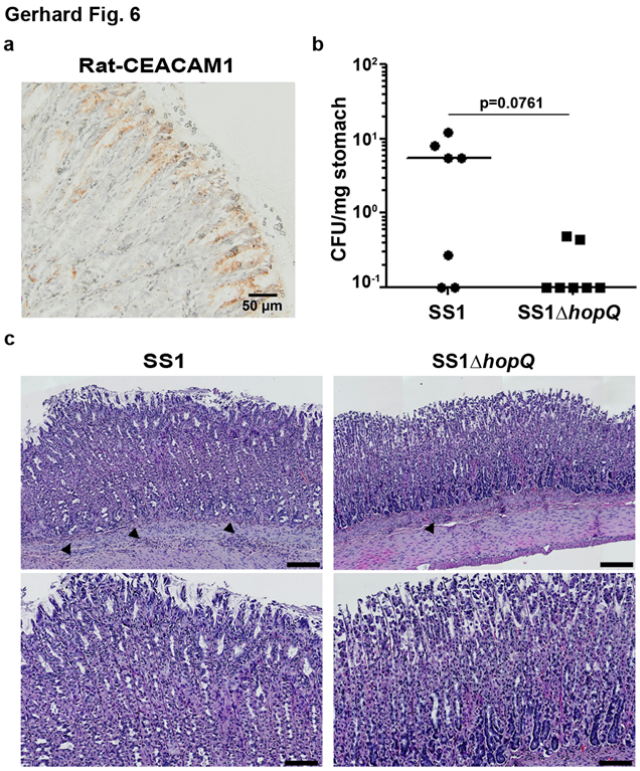


Figure 6 *H. pylori* colonization of rat stomach depends on HopQ. (a) CEACAM1 expression in rat stomach. (b) *H. pylori* colony forming units (CFU) per mg stomach of male Sprague dawley rats after 6 weeks infection. Horizontal bars indicate medians. Mann-Whitney U test. (c) Hematoxylin/eosin staining of infected rat stomachs. Representative images of same stomach regions are shown. Scale bar 100µm (upper panels) and 200µm (lower panels). Arrows denote inflammatory cells.

H. pylori adhesin HopQ engages in a virulence-enhancing interaction with human CEACAMs

Anahita Javaheri[1,15,‡], Tobias Kruse[2,‡], Kristof Moonens[3,4,‡], Ayla Debraekeleer[3,4], Raquel Mejías-Luque[1,15], Isabell Asche[5], Nicole Tegtmeyer[5], Behnam Kalali[1,2], Nina C. Bach[6], Stephan A. Sieber[6], Darryl J. Hill[7], Verena Königer[8], Christof R. Hauck[9], Roman Moskalenko[10], Rainer Haas[8], Dirk H. Busch[1], Esther Klaile[11,12], Hortense Slevogt[11], Alexej Schmidt[13,14], Steffen Backert[5], Han Remaut[3,4,‡], Bernhard B. Singer[12‡] and Markus Gerhard[1,2,15‡*]

Affiliations:

[1]Institute for Medical Microbiology, Immunology and Hygiene; Technische Universität München; Munich, 81675, Germany,

[2]Imevax GmbH, 81675 Munich

[3]Structural and Molecular Microbiology, Structural Biology Research Center, VIB, Pleinlaan 2, 1050 Brussels, Belgium

[4]Structural Biology Brussels, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussels, Belgium

[5]Friedrich Alexander University Erlangen, Department of Biology, Division of Microbiology, Erlangen, Germany

[6]Center for Integrated Protein Science Munich, Department Chemie, Institute of Advanced Studies, Technische Universität München, 85747 Garching, Germany

[7]School of Cellular & Molecular Medicine, University of Bristol, BS8 ITD, Bristol, UK

[8]Max von Pettenkofer-Institut für Hygiene und Medizinische Mikrobiologie, Department of Bacteriology, Ludwig-Maximilians-Universität, D-80336 Munich, Germany

[9]Lehrstuhl für Zellbiologie, Universität Konstanz, Konstanz, Germany

[10]Department of Pathology, Sumy State University, Sumy 40000, Ukraine

[11]Septomics Research Centre, Jena University Hospital, 07745 Jena, Germany.

[12]Center for Sepsis Control and Care (CSCC), Jena University Hospital, 07747 Jena, Germany

[13]Institute of Anatomy, Medical Faculty, University Duisburg-Essen, 45122 Essen, Germany

[14]Department of Medical Biosciences, Pathology, Umeå University, SE-901 85 Umeå, Sweden

[15]German Center for Infection Research, Partner Site Munich, Munich, Germany

*Correspondence to: markus.gerhard@tum.de

‡ These authors contributed equally to this work

Summary: *Helicobacter pylori* specifically colonizes the human gastric epithelium and is the major causative agent for ulcer disease and gastric cancer development. Here we identified members of the carcinoembryonic antigen-related cell adhesion molecule (CEACAM) family as novel receptors of *H. pylori* and show that HopQ is the surface-exposed adhesin that specifically binds human CEACAM1, CEACAM3, CEACAM5 and CEACAM6. HopQ - CEACAM binding is glycan-independent and targeted to the N-domain. *H. pylori* binding induces CEACAM1 mediated signaling, and the HopQ-CEACAM1 interaction enables translocation of the virulence factor CagA into host cells, and enhances the release of pro-inflammatory mediators such as interleukin-8. Based on the crystal structure of HopQ, we found that a β-hairpin insertion (HopQ-ID) in HopQ's extracellular 3+4 helix bundle domain is important for CEACAM binding. A peptide derived from this domain competitively inhibits HopQ-mediated activation of the Cag virulence pathway, as genetic or antibody-mediated abrogation of the HopQ function shows. Together, our data imply the HopQ-CEACAM1 interaction as potentially promising novel therapeutic target to combat *H. pylori*-associated diseases.

Helicobacter pylori (*H. pylori*) is one of the most prevalent human pathogens, colonizing half of the world's population. Chronic inflammation elicited by this bacterium is the main cause of gastric cancer[1]. During co-evolution with it's human host over more than 60.000 years[2], the bacterium has acquired numerous adaptations for the long-term survival within its unique niche, the stomach. This includes the ability to buffer the extreme acidity of this environment, the interference with cellular signaling pathways, the evasion of the human immune response and a strong adhesive property to host cells[3]. Specifically, *H. pylori* persistence is facilitated by the binding of BabA and SabA adhesins to the human blood group antigen Leb and the sLex antigen, respectively[4-6]. However, adhesion to blood group antigens is not universal, is dynamically regulated during the course of infection and can also be turned off [7]. We observed that *H. pylori* was capable of binding to human gastric epithelium of non-secretors. Therefore, we hypothesized that the bacterium might be able to interact with other cell surface receptors to ensure persistent colonization.

We here show that the *H. pylori* adhesin HopQ specifically interacts with human carcinoembryonic antigen-related cell adhesion molecules (CEACAMs). CEACAMs embrace a group of immunoglobulin superfamily-related glycoproteins with a wide tissue distribution. CEACAM1 can be expressed in leukocytes, endothelial and epithelial cells, CEACAM3 and CEACAM8 in granulocytes, CEACAM5 and CEACAM7 in epithelial cells and CEACAM6 in epithelia and granulocytes. In epithelial cells, transmembrane anchored CEACAM1 as well as glycosylphosphatidylinositol-linked CEACAM5, CEACAM6 and CEACAM7 localize to the apical membrane[8]. CEACAMs modulate diverse cellular functions such as cell adhesion, differentiation, proliferation, and cell survival. Some CEACAMs were recognized as valuable tumor markers due to their enlarged expression in the malignant tissue and increased sera level[9]. In recent years, CEACAMs have also emerged as immunomodulatory mediators[10]. Interestingly, in humans, several CEACAMs have been found to specifically interact with bacteria such as *Neisseria, Haemophilus influenzae*, *Moraxella catarrhalis*, and *Escherichia coli* [11].

H. pylori binds to CEACAMs expressed in human stomach

Based on the observation that *H. pylori* efficiently colonizes individuals in the absence of Lewis blood group antigens[12] on the one hand, and the increased expression of members of the carcinoembryonic antigen-related cell adhesion molecule family (CEACAMs) in gastric tumors, we hypothesized that *H. pylori* may employ CEACAMs as receptors. Using pull down and flow cytometric approaches we found a robust interaction of the *H. pylori* strain

G27 with recombinant human CEACAM1-Fc (Fig. 1a), comparable to that of *Moraxella catarrhalis* (Extended Data Fig. 1a and b). As negative control, *Moraxella lacunata* did not bind to human CEACAM1, nor did *Campylobacter jejuni,* a pathogen closely related to *H. pylori* (Extended Data Fig. 1a and b). When testing for CEACAM specificity, we observed a clear interaction of *H. pylori* also with CEACAM3, 5 and 6, but not with CEACAM8 (Fig.1b and Extended Data Fig. 1c and d). Importantly, all *H. pylori* strains tested bound to these CEACAMs (Extended Data Fig.1f and g) including well-characterized reference strains (26695, J99) and the mouse-adapted strain SS1. However, binding strength differed among strains, with some preferentially binding to CEACAM1, and others to CEACAM5 and/or CEACAM6 (Extended Data Fig. 1f and g). We then analyzed the expression profiles of CEACAM1, CEACAM5 and CEACAM6 in normal and inflamed human stomach tissues and gastric cancer. If at all low levels of CEACAM1 and CEACAM5 were expressed at the apical side of epithelial cells, and their expression, as well as that of CEACAM6, was up-regulated upon gastritis and in gastric tumors (Fig. 1c and Extended Data Fig. 1e). During infection, *H. pylori*-induced responses may thus lead to increased expression of its CEACAM-receptors.

Adhesins from other bacteria were shown to specifically bind to the N-domain of human CEACAM1[13,14]. Similarly, we found that lack of the CEACAM1 N-domain abolished *H. pylori* binding completely (Fig. 1d). While for the interaction of *Neisseria meningitidis* with CEACAM1 the N-domain was necessary but not sufficient for binding[15], we observed binding of *H. pylori* to all tested CEACAM1 isoforms containing the N-domain, as well as to the N-domain alone (Fig. 1e). However, binding to the N-domain alone was weaker than to the N-A1-B CEACAM1 variant, which bound less than the N-A1-B-A2 variant (Fig. 1e and Extended Data Fig.1j), suggesting that these domains stabilize the CEACAM1-*H.pylori* interaction. Comparison of the respective N-domains indicated several residues conserved in CEACAM1, 5, and 6 but not in CEACAM8 (Extended Data Fig. 1h).

Species specificity of *Helicobacter* – CEACAM interaction

Although, murine and Mongolian gerbil models are routinely used to study gastric infection with *H. pylori*, the bacterium has been described so far to be naturally transmitted to only humans and non-human primates. Although CEACAMs are found in most mammalian species, and have a high degree of conservation, we found *H. pylori* to bind selectively to human, but not to mouse, bovine or canine CEACAM1 orthologues (Fig. 2a). However, we were surprised to find a strong interaction of *H. pylori* with rat-CEACAM1 (Fig. 2b and d). This interaction was also mediated through the N-domain of rat-CEACAM1 (Fig. 2c and d).

To substantiate these findings, we transfected human, mouse or rat-CEACAM1 into CHO cells, to which *H. pylori* does not adhere otherwise. Using confocal laser scanning microscopy, we observed *de novo* adhesion of *H. pylori* to CHO cells expressing human and rat, but not mouse CEACAM1 (Fig. 2e), which could be confirmed by pull down and Western blotting of lysates from transfected cells (Fig. 2f and Extended Data Fig. 2d). This finding makes *H. pylori* the first pathogen for which its CEACAM binding is not restricted to one species. Comparing the protein sequences of the CEACAM1-N domains, several amino acids conserved in human and rat differ in mouse (i.e. asn10, glu26, asn42, tyr48, pro59, thr66, asn77, val79, val89, ile90, glu103, tyr108) (Extended Data Fig. 2a). In addition, our findings of the lack of binding to mouse CEACAM1 may explain the differences seen in pathology between infected mice and humans[16].

The genus *Helicobacter* comprises several other spp. i.e. *H. felis, suis, and bizzozeronii* as well as the human pathogenic *H. bilis* and *H. heilmannii*. When assessing the interaction of these *Helicobacters* with human CEACAMs, only *H. bilis* bound to human CEACAM1, 5 and 6 (Extended Data Fig.2b and c). As *H. pylori*, *H. bilis* interacted with the N-domain of hu-CEACAM1 (Extended Data Fig.2b and c). This interaction may explain how *H. bilis* manages to colonize human bile ducts, where high levels of constitutively expressed CEACAM1 are present.

HopQ is the *Helicobacter* adhesin interacting with CEACAMs

In order to identify the CEACAM-binding partner in *Helicobacter,* we initially screened a number of *Helicobacter* mutants devoid of defined virulence factors that have been shown to be implicated in various modes of host cell interaction (BabA, SabA, AlpA/B, VacA, gGT, urease and the *cag*PAI)[5,6,17]. All of these mutants still bound to hu-CEACAM1 (Fig. 3a). Therefore we established an immunoprecipitation approach (Extended Data Fig. 3a) using *H. pylori* lysate and recombinant hu-CEACAM1-Fc coupled to protein G. Mass spectrometric analysis of the co-precipitate identified two highly conserved *H. pylori* outer membrane proteins as candidate CEACAM1 adhesins: HopQ and HopZ (Fig. 3b). Unlike a *hopZ* mutant, a *hopQ* deletion mutant was devoid of CEACAM1 binding (Fig. 3c). Importantly, the *hopQ* mutant was also unable to bind to CEACAM5 and 6 (Fig.3c).

Next we tested the binding of recombinant HopQ to different gastric cancer cell lines and found that HopQ interacted with AGS and MKN45 both endogenously expressing CEACAMs (Extended Data Fig.3b). HopQ did not bind to the CEACAM negative cell line MKN28. Utilizing our CHO transfectants, we found that the recombinant HopQ interacted

preferentially with CEACAM1 and 5, and to lesser extent to CEACAM3 and 6. No binding was observed to CHO cells expressing either CEACAM4, 7, or 8 (Extended Data Fig. 3c).

HopQ is a member of a *H. pylori*-specific family of outer membrane proteins, and shows no significant homology to other CEACAM-binding adhesins from other Gram-negative bacteria, i.e. Opa proteins or UspA1 from *Neisseria meningitidis* and *Neisseria gonorrhoeae* or *Moraxella catarrhalis*, respectively, and is therefore a novel bacterial factor hijacking CEACAMs. Like Opa and UspA1 [13,14], HopQ targets the N-terminal domain in CEACAMs, an interaction we found to require folded protein (see below) and was dependent on CEACAM sequence, resulting in specificity for human CEACAM1, 3, 5 and 6. The *H. pylori hopQ* gene (*omp27*; HP1177 in the *H. pylori* reference strain 26695) exhibits genetic diversity that represents two allelic families[18], type-I and type-II (Extended Data Fig. 3d), of which the type-I allele is found more frequently in *cag*(+)/s1-*vacA* type strains. Both alleles share 75 to 80% nucleotide sequences and exhibit a homology of 70% at the amino acid level[18]. Importantly, *hopQ* genotype shows a geographic variation, with the *hopQ* type-I alleles more prevalent in Asian compared to Western strains; and was also found to correlate with strain virulence, with type-I alleles associated with higher inflammation and gastric atrophy[19].

Structure and binding properties of the HopQ adhesin domain

HopQ belongs to a paralogous family of *H. pylori* outer membrane proteins (Hop's), to which also the blood group antigen binding adhesins BabA and SabA belong[5,6,17,20].To gain insight into its structure-function relationship we determined the binding properties and X-ray structure of a HopQ fragment corresponding to its predicted extracellular domain (residues 17-444 of the mature protein; HopQ^{AD}; Fig. 4a). HopQ^{AD} showed strong, dose dependent binding to the N-terminal domain of human CEACAM1 (C1ND; residues 35-142) in ELISA (Fig. 4b) and isothermal titration calorimetry (ITC) revealed a 1:1 stoichiometry with a dissociation constant of 296±40 nM (Extended Data Fig. 4a). The HopQ^{AD} X-ray structure shows that, like BabA and SabA, the HopQ ectodomain adopts a 3+4-helix bundle topology, though lacks the extended coiled-coil "stem" domain that connects the ectodomain to the transmembrane region (Fig. 4a and Extended Data Fig.4d). In BabA, the carbohydrate binding site resides fully in a 4-stranded β-domain that is inserted between helices 4 and 5[21] (Extended Data Fig.4d). In HopQ, a 2-stranded β-hairpin is found in this position (residues 180-218). Removal of the β-hairpin resulted in a soluble protein that showed a ~10 fold reduction of CEACAM1 binding affinity (Fig. 4b and Extended Data Fig. 4c), indicating that although the

HopQ insertion domain is implicated in binding, it does not comprise the full binding site as found in BabA (Fig. 4b).

The hitherto characterized Hop adhesins are lectins[5,6,17,22]. Instead, *H. pylori* was seen to retain binding to CEACAM1 upon enzymatic deglycosylation, and Far Western analysis revealed that $HopQ^{AD}$ specifically bound folded, but not denatured C1ND (Fig. 4c), suggesting HopQ-CEACAM binding relies on protein-protein rather than glycan-dependent interactions. Indeed, ITC binding profiles of $HopQ^{AD}$ titrated with non-glycosylated *E. coli* expressed C1ND (Ec-C1ND) revealed an equimolar interaction with a dissociation constant of 417±48 nM (Extended Data Fig. 4b), showing that CEACAM N-glycosylation only provides a minor stabilizing contribution to the HopQ-CEACAM interaction. To further map the HopQ binding site, we pre-incubated CEACAM1 with the *M. catarrhalis* adhesin UspA1, and found that this prevented binding by *H. pylori* (Fig. 4d), suggesting that both adhesins have overlapping binding epitopes. In further support, mutation of CEACAM1 residues Y34 or I91 within the UspA1 binding epitope reduced or nearly abrogated CEACAM1 binding by *H. pylori* (Fig. 4e). Interestingly, I91 is conserved in rat but mutated to T in mouse CEACAM1, possibly explaining the observed species specificity in HopQ binding (Extended Data Fig. 2a, see above).

HopQ – CEACAM1 interaction triggers cell responses

Available animal models only partially replicate the *H. pylori* pathogenesis observed in its human host and mouse CEACAMs did not support HopQ binding. Therefore, to further investigate how HopQ may influence adhesion and cellular responses, we sought to establish cellular pathogenesis models in which the HopQ-CEACAM mediated adhesion could be analyzed. According to Singer et al.[23], we characterized various gastric cell lines typically employed for *H. pylori in vitro* experiments regarding their expression of CEACAMs, and observed that MKN45, KatoIII and AGS did express CEACAM1, CEACAM5 and CEACAM6, whereas MKN28 showed no presence of CEACAMs (Extended Data Fig.5a and b). In parallel, CHO cells were stably transfected with CEACAM1-L (containing the immunoreceptor tyrosine-based inhibition motif (ITIM). Upon infection with *H. pylori* wild-type strain P12 and its isogenic *hopQ* deletion mutant, we observed a significantly reduced adherence to CHO-CEACAM1-L, MKN45 and AGS cells when hopQ was not present, while strains deficient in the adhesins BabA and SabA showed only slightly reduced adhesion (Fig. 5a and Extended Data Fig.5c). HopQ binding was also studied in human gastric biopsies from *H. pylori* infected individuals. Here, we detected that HopQ bound to the apical side human

gastric epithelium and co-localized with CEACAM in biopsies from *H. pylori* infected individuals (Fig. 5b and Extended Data Fig. 5d), while no binding was observed in CEACAM1 negative samples from normal stomach (not shown). In CHO-CEACAM1-L cells, we observed tyrosine-phosphorylation of the CEACAM1 ITIM domain upon exposure to *H. pylori*, which was apparent within 5 minutes, and was maintained for up to 1 hour (Fig.5c). Phosphorylation of the CEACAM1 ITIM domain is a well-known initial event triggering SHP1/2 recruitment inducing downstream signaling cascades[24,25]. Contact-dependent signaling through CEACAMs is a common means of modulating immune responses related to infection, inflammation and cancer[10], and these immune-dampening cascades likely reflect the multiple independent emergence of non-homologous CEACAM-interacting proteins in diverse mucosal Gram-negative pathogens including *Neisseria, Haemophilus, Escherichia, Salmonella, Moraxella* sp.[13,14]. For *H. pylori*, interaction with human CEACAM1 through HopQ may represent a critical parameter for immuno-modulatory signaling during colonization and chronic infection of man.

Additionally, *hopQ* mutant *H. pylori* strains showed an almost complete loss of *cag*PAI-dependent CagA translocation (Fig. 5d) and strongly reduced IL-8 induction (Fig.5e), while loss of other known adhesins had no effect on CagA delivery (Extended Data Fig.5e and f). This is in line with a previous study showing that in AGS gastric cancer cells, a *hop*Q mutant *H. pylori* strain exhibited reduced ability to activate NF-κB and altered translocation of CagA[26]. In contrast to our findings, Belogolova et al. did not observe reduced adherence of a *hopQ* mutant *H. pylori* P12 strain, which could be due to the observed growth dependent expression of CEACAMs in these cells.

To corroborate our data in an independent model and compensate for potential clonal effects in stably transfected cells, we transiently transfected HEK293 cells with human CEACAM (1-L,3,4,5,6,7,8) expression plasmids. Infection of these cells confirmed the defect in CagA translocation observed in CHO-CEACAM1-L cells, which was restored upon complementation of the *hopQ* mutant strain (P12Δ*hopQhopQ*$^{+}$) (Fig.5f and Extended Data Fig.5g). Also, cellular elongation, the so called "hummingbird phenotype", was significantly reduced upon deletion of *hopQ* (Fig. 5g and h). Further, we observed that *H. pylori* modulates important host transcription factors such as Myc or STAT3, in a *hopQ*-dependent fashion (Extended Data Fig. 5h). Our results reveal that HopQ-CEACAM binding leads to direct and indirect alterations in host cell signaling cascades, and start to shed light on these HopQ-associated virulence landscapes. Given the importance of these signaling events for gastric carcinogenesis, we explored if the CEACAM-HopQ interaction could be targeted in order to

prevent CagA translocation and downstream effects. Indeed, incubation of the cells with an α-CEACAM1 antibody, α-HopQ antiserum or a HopQ-derived peptide corresponding to the Hop-ID (aa 189-220) reduced CagA translocation in a dose dependent manner (Fig. 5i-k), but not corresponding controls (Extended Data Fig. 5h). These data demonstrate that the HopQ-CEACAM1 interaction is necessary for successful translocation of the oncoprotein CagA into epithelial cells as well as modulation of inflammatory signaling, and that interference with this interaction can prevent CagA translocation, giving an indication of the translational potential of HopQ targeting for *H. pylori* vaccination or immunotherapy.

Deletion of *hopQ* abrogates colonization in a rat model of *H. pylori* infection

As we have found binding of HopQ to human and rat, but not to mouse CEACAM, we finally determined the role of HopQ *in vivo,* using a rat model of *H. pylori* infection. Having observed that CEACAM1 was expressed in normal rat stomach (Fig. 6a and Extended Data Fig. 6b), we infected rats with the mouse adapted strain SS1, able to bind human and rat CEACAM1 (Extended Data Fig. 6a). While the wilt type SS1 was able to efficiently colonize rats, albeit at lower levels compared to the mouse, (Fig. 6b) , the *hopQ* deficient SS1 strain was not able to colonize rats at detectable levels, and could not induce an inflammatory response in comparison to the wild type SS1 strain (Fig. 6b and c). Therefore, in this model, HopQ seems also to serve as an important factor to mediate *H. pylori* colonization. While infection of rats with *H. pylori* has been described[27], our finding may allow the establishment of an animal model for studying *H. pylori* infection that better replicates the prevailing virulence pathways.

Discussion

The here identified CEACAM-binding property provides *H. pylori* a means of epithelial adherence in addition to the Lewis antigens used by the BabA and SabA adhesins[5,6,17]. While over-expression of CEACAMs in gastrointestinal tumors is well described, their up-regulation during *H. pylori*-induced inflammation in the stomach has not been reported so far, suggesting the pathogen has the ability to shape its own adhesive niche. A similar phenomenon has also been observed for the inflammation-induced up-regulation of sialylated antigens that form the receptors for the SabA adhesin[6]. A plausible route to CEACAM modulation is through the transcription factors NF-κB and AP1, both of which are induced during *H. pylori* infection[28] and are known to regulate CEACAM expression[29]. Though HopQ-dependent adherence may appear redundant to that of other adhesins like BabA, SabA or LabA, HopQ specializes on human CEACAMs and is required for *cag*PAI functionality. From the perspective of host-pathogen (i.e. human-*H. pylori*) co-evolution, the primary function of HopQ may lie in immune-modulation through CEACAM binding, and HopQ's indirect effects on other virulence cascades elicited by *H. pylori* such as that induced by increased CagA delivery may not have been initially "intended". The *cag*PAI was acquired by ancestral *H. pylori* in a single event that occurred before modern humans migrated out of East Africa around 58,000 years ago[30]. Thus, it is likely that the employment of CEACAM1 ligation by *H. pylori* occurred much earlier to support colonization and to modulate immune responses. This assumption is supported by the fact that all fully sequenced *H. pylori* strains bear *hop*Q (Extended Data Fig.3d), indicating that this is an essential outer membrane protein of *H. pylori*. Upon occurrence of type-I *H. pylori* strains by *cag*PAI acquisition more than 60,000 years ago[30] this ancient survival strategy was further implemented into a mechanism supporting pathogenicity, and thus may have contributed to the switch from commensal to pathogenic *H. pylori*[31]. Pathogenicity might even be further aggravated by our observation that CEACAMs are strongly up-regulated during gastritis, which further potentiates binding of *H. pylori* to epithelial cells and specifically facilitates CagA/cagPAI interaction with the host cells.

Taken together, the finding that *H. pylori* employs CEACAMs not only for bacterial adherence but also to induce cellular signaling may lead to a better understanding of the pathogenic mechanisms of these bacteria and might lead to novel therapeutic approaches to more effectively combat this highly prevalent infection and the associated gastric pathology.

Materials and Methods

Bacteria and bacterial growth conditions

The *H. pylori* strains G27[32], PMSS1[33], SS1[34], J99 (ATCC, 700824), 2808[35], 26695 (ATCC, 70039), TX30[36], 60190[37], P12[38], NCTC11637 (ATCC, 43504), Ka89 and *H. bilis* (ATCC43879) were grown on Wilkins–Chalgren blood agar plates under microaerobic conditions (10% CO2, 5% O2, 8.5% N2, and 37°C). *H. suis*[39] and *H. heilmannii*[40] were grown on Brucella agar and *H. felis* (ATCC 49179) and *H. bizzozeronii*[41] on brain-heart infusion (BHI) agar supplemented with 10% horse blood. *Moraxella catarrhalis* (ATCC, 25238) provided by C. R. Hauck (Konstanz Research School Chemical Biology, University of Konstanz, Germany), *Moraxella Lacunata* (ATCC 17967) and *Campylobacter jejuni* (ATCC, 33560) were cultured on brain–heart infusion (BHI) agar supplemented with 5% heated horse blood overnight at 37°C in a CO_2 incubator. The generation of an isogenic Δ*hopQ* mutant has been done by replacement of the entire gene by a chloramphenicol resistance cassette. For genetic complementation of *hopQ*, the 1,926 bp gene fragment of *H. pylori* strain P12 was amplified by PCR. This fragment was cloned into the complementation vector pSB1001 using the AphA3 cassette for selection. This fusion construct was introduced in the plasticity region of strain P12Δ*hopQ* (between ORFs HP0999 and HP1000) using a strategy as described[42].

Production of CEACAM proteins

The cDNA, which encodes the extracellular domains of human CEACAM1-Fc (consisting of N-A1-B1-A2 domains), human CEACAM1dN-Fc (consisting of A1-B1-A2, lacking the first 143 amino acids of the N-terminal IgV-like domain), rat CEACAM1-Fc (consisting of N-A1-B1-A2), rat CEACAM1dN-Fc (consisting of A1-B1-A2), human CEACAM3-Fc (consisting of N), human CEACAM6-Fc (consisting of N-A-B), human CEACAM8-Fc (consisting of N-A-B), respectively, were fused to a human heavy chain Fc-domain and cloned into the pcDNA3.1(+) expression vector (Invitrogen, San Diego, CA), sequenced and stably transfected into HEK293 (ATCC CRL-1573)cells as described[43]. The Fc chimeric CEACAM-Fc proteins were accumulated in serum-free Pro293s-CDM medium (Lonza) and were recovered by Protein A/G-Sepharose affinity Chromatography (Pierce). Proteins were analyzed by SDS-PAGE and stained by Coomassie blue demonstrating an equal amount and integrity of the produced fusion proteins (Extended Data Fig. 1i). Recombinant-human CEACAM5-Fc was ordered from Sino Biological Inc. The GFP-tagged CEACAMs (human-CEACAM1 and its variants, mouse-CEACAM1, bovine-CEACAM1 and canine-CEACAM1)

were provided by Dr. C. R. Hauck (University Konstanz, Germany). For production of the recombinant human CEACAM1 N-Domain (C1ND), the annotated domain (residues 35-142 of CEACAM1, Uniprot ID: P13688) was first backtranslated using the Gene Optimizer® (LifeTechnologies) and the leader sequence of the Igk-chain as well as a C-terminal Strep-Tag II was added. The gene was synthesized and seamlessly cloned into pCDNA3.4-TOPO (LifeTechnologies). Protein was produced in a 2 L culture of Expi293 cells according to the Expi293 expression system instructions (LifeTechnologies). The resulting supernatant was concentrated and diafiltered against ten volumes of 1x SAC buffer (100 mM Tris-HCl, 140 mM NaCl, 1 mM EDTA, pH 8.0) by crossflow-filtration, using a Hydrosart 5 kDa molecular-weight cutoff membrane (Sartorius). The retentate was loaded onto a StrepTrap HP column (GE Healthcare) and eluted with 1x SAC supplemented with 2.5 mM D-Desthiobiotin (IBA). The protein was stored at +4°C.

For the bacterial expression of the C1ND (Ec-C1ND) the amino acid sequence (residues 35-142 of CEACAM1, Uniprot ID: P13688) was codon optimized for expression in *E. coli*, synthesized by GeneArt *de novo* gene synthesis (Life Technologies), and cloned with a C-terminal His6 tag in the pDEST™14 vector using Gateway technology (Invitrogen). *E. coli* C43(DE3) cells were transformed with the resulting construct and grown in LB supplemented with 100 µg/mL ampicillin at 37°C while shaking. At OD_{600}=1 Ec-C1ND expression was induced with 1 mM IPTG overnight at 30°C. Cells were collected by centrifugation at 6.238 *g* for 15 minutes at 4°C and resuspended in 50mM Tris-HCl pH 7.4, 500 mM NaCl (4 mL/g wet cells) supplemented with 5 µM leupeptin and 1 mM AEBSF, 100 µg/mL lysozyme, and 20 µg/mL DNase I. Subsequently cells were lysed by a single passage in a Constant System Cell Cracker at 20 kPsi at 4 °C and debris was removed by centrifugation at 48.400 *g* for 40 minutes. The cytoplasmic extract was filtrated through a 0.45 µm pore filter and loaded on a 5 mL pre-packed Ni-NTA column (GE Healthcare) equilibrated with buffer A (50 mM Tris-HCl pH 7.4, 500 mM NaCl and 20 mM imidazole). The column was then washed with 40 bed volumes of buffer A and bound proteins were eluted with a linear gradient of 0-75 % buffer B (50 mM Tris-HCl pH 7.4, 500 mM NaCl and 500 mM imidazole). Fractions containing Ec-C1ND, as determined by SDS-PAGE, were pooled and concentrated in a 10 kDa MW cutoff spin concentrator to a final volume of 5 ml. To remove minor protein contaminants, the concentrated sample was injected onto the Hi-Prep™ 26/60 Sephacryl S-100 HR column (GE Healthcare) pre-equilibrated with a buffer containing 50 mM Tris-HCl pH 8.0, 150 mM NaCl. Fractions containing the Ec-C1ND complex were pooled and concentrated using a 10 kDa MW cutoff spin concentrator.

HopQ^{AD} and HopQ^{AD}ΔID cloning, production and purification

In order to obtain a soluble HopQ fragment, the HopQ gene from the *H. pylori* G27 strain (accession No. CP001173 Region: 1228696..1230621) HopQ fragment ranging from residues 37 – 463 was produced (residues 17-444 of the mature protein), thus removing the N-terminal β-strand and signal peptide, as well as the C-terminal β-domain expected to represent the TM domain. In HopQ^{AD}ΔID, the amino acids 184-212 of the mature protein were replaced by two glycines (Extended Data Fig.f). DNA coding sequences corresponding to the HopQ type I fragments was PCR-amplified from *H. pylori* G27 genomic DNA using primers (forward: GTTTAACTTTAAGAAGGAGATATACAAATGGCGGTTCAAAAAGTGAAAAACGC; reverse: TCAAGCTTATTAATGATGATGATGATGGTGGGCGCCGTTATTCGTGGTTG), containing 30bp overlap to the flanking target vector sequences of pPRkana-1, a derivative of pPR-IBA 1 (IBA GmbH) with the ampicillin resistance cassette replaced by the kanamycin resistance cassette, under a T7 promotor. In parallel, the vector was PCR-amplified using primers (forward: CACCATCATCATCATCATTAATAAGCTTGATCCGGCTGCTAAC ; reverse: GTTTAACTTTAAGAAGGAGATATACAAATG) as provided in table 1, using the same overlapping sequences in reversed orientation. The forward primer additionally carried the sequence for a 6x His-tag. The amplicons were seamlessly cloned using Gibson Assembly (New England Biolabs GmbH). Based on codon optimized HopQ^{AD} plasmid, the HopQ^{AD}ΔID constructs were cloned. The plasmids were amplified by 5' phosphorylated primers (forward: GGTGACGCTCAGAACCTGCTGAC; reverse: ACCACCTTTAGAGTTCAGCGGAG) replacing the ID region by two glycines, *Dpn*I (NEB) digested and blunt-end ligated by T4 ligase (NEB).

Escherichia coli BL21(DE3) cells (NEB GmbH) were transformed with the pPRkana-1 constructs, grown at 37°C with 275 rpm on auto-inducing terrific broth (TRB) according to Studier[44], supplemented with 2 mM $MgSO_4$, 100 mg/L Kanamycin-Sulfate (Carl Roth GmbH + Co. KG), 0.2 g/L PPG2000 (Sigma-Aldrich) and 0.2% w/v Lactose-monohydrate (Sigma-Aldrich), until an OD of 1-2 was reached. Afterwards, the temperature was lowered to 25°C and auto-induced overnight, reaching a final OD of 10-15 the following morning. Cells were harvested by centrifugation at 6000 *g* for 15 min at 4 °C using a SLA-3000 rotor in a Sorvall RC-6 Plus centrifuge (Thermo Fischer). Prior to cell disruption, cells were resuspended in 10 mL cold NiNTA buffer A (500 mM NaCl, 100 mM Tris-HCl, 25 mM Imidazole, pH 7.4) per gram of biological wet weight (BWW), supplemented with 0.1 mM AEBSF-HCl, 150 U/g

BWW DNase I and 5 mM $MgCl_2$and dispersed with an Ultra-Turrax T25 digital (IKA GmbH + Co. KG). Cell disruption was performed by high-pressure homogenization with a PANDA2000 (GEA NiroSoavi) at 800-1200 bar in 3 passages at 4 °C. The cell lysate was clarified by centrifugation at 25000 *g* for 30 min at 4 °C in a SLA-1500 rotor and remaining particles removed by filtration through a 0.2 µM filter.

HopQ fragments were purified by consecutive nickel affinity and size exclusion chromatography. Briefly, the clarified cell lysate was loaded onto a 5 mL pre-packed Ni-NTA HisTrap FF crude column (GE Healthcare) pre-equilibrated with buffer A, washed with ten column volumes (CV) of buffer A and the bound protein eluted with a 15 CV linear gradient to 75% NiNTA buffer B (500 mM NaCl, 100 mM Tris-HCl, 500 mM Imidazole, pH 7.4). Eluted peak fractions were collected, pooled and concentrated to a final concentration of 8-10 mg ml^{-1} using a 10 kDa molecular-weight cutoff spin concentrator. Subsequently, 5 mL of the concentrated protein were loaded onto a HiLoad 16/600 Superdex 75 pg column (GE Healthcare) pre-equilibrated with Buffer C (5 mM Tris-HCl, 140 mM NaCl, pH 7.3) and eluted at 1 mL/min. Finally, only protein corresponding to the monomer-peak was pooled and stored at +4 °C prior to crystallization. For analyzing the multimerization state of $HopQ^{AD}$, SEC was performed on a Superdex 200 10/300 GL (GE Healthcare) with 24 mL bed volume. The column was pre-equilibrated with Buffer C and subsequently, 25 µg protein injected and separated with a flow rate of 0.5 mL/min.

The HopQ interaction domain (HopQ-ID) representing peptide was HA-tagged, synthesized (EKLEAHVTTSKYQQDNQTKTTTSVIDTTNYPYDVPDYA) and HPLC purified (Peptide Specialty Laboratories, Heidelberg, Germany). For cellular assays, the lyophilized peptide was dissolved in sterile PBS to a concentration of 1 mM and dialysed with a 0.1-0.5 kDa molecular-weight cutoff membrane against PBS to remove remaining TFA. The peptide solution was stored at -20 °C until further use.

Detection of the HopQ-CEACAM interaction by ELISA

For detection of the interaction between CEACAM and $HopQ^{AD}$, recombinant C1ND (1 µg/mL) in PBS was coated over night at 4 °C onto a 96-well immunoplate (Nunc MaxiSorb). Wells were blocked with SmartBlock (Candor) for 2 h at RT. Subsequently, HopQ fragments were added in a fivefold series dilution ranging from 10 µg/mL to 0.05 ng/mL for 2h at room temperature. Next, anα-6xHis-HRP conjugate (clone 3D5, LifeTechnologies) was diluted 1:5000 and incubated for 1h at room temperature. For detection, 1-Step™ Ultra TMB-ELISA Substrate Solution (LifeTechnologies) was used and the enzymatic reaction was stopped with

2 N H_2SO_4. Washing (3-5x) in between incubation steps was carried out with PBS / 0.05% Tween20.

Isothermal titration calorimetry

ITC measurements were performed on a MicroCal iTC200 calorimeter (Malvern). 25 µM C1ND or EcC1ND were loaded into the cell of the calorimeter and 250 µM $HopQ^{AD}$ type I was loaded in the syringe. All measurements were done at 25°C, with a stirring speed of 600 rpm and performed in 20 mM HEPES buffer (pH 7.4), 150 mM NaCl, 5% (v/v) glycerol and 0.05% (v/v) Tween-20. Binding data were analyzed using the MicroCal LLC ITC200 software.

SDS-PAGE and native-PAGE for Western blot

CEACAM was separated with both SDS-PAGE and native-PAGE (resp. on 15% and 7.5% polyacrylamide gels) in ice-cold 25 mM Tris-HCl, 250 mM glycine buffer. Subsequently samples were transferred to PVDF-membranes by wet blotting at 25 V during 60 minutes in ice-cold transfer buffer (25 mM Tris-HCl, 250 mM glycine and 20% methanol). Membranes were blocked during one hour in 10% milk powder (MP), 1x PBS and 0.005% Tween-20. Both membranes were washed and incubated together in 5% MP, 1x PBS, 0.005% Tween-20 in presence of 2 µM $HopQ^{AD}$ type I for one hour to allow complex formation between $HopQ^{AD}$ I and CEACAM. After a washing step the C-terminal His-tag of HopQ (CEACAM is strep tagged) was detected by adding consecutively mouse α-His (AbDSerotec) and goat α-mouse antibody (Sigma-Aldrich) during respectively one hour and 30 minutes in 5% MP, 1x PBS, 0.005% Tween-20. After a washing step the blot was developed by adding BCIP/NBT substrate (5-bromo-4-chloro-3-indolyl-phosphate/nitro blue tetrazolium) (Roche) in developing buffer (10 mM Tris-HCl pH 9.5, 100 mM NaCl, 50 mM $MgCl_2$).

Bacterial pull down

Bacteria were grown overnight on WC dent agar plates. Bacteria were scraped from plates, suspended in PBS, and colony forming units (cfu) were estimated by optical density 600 readings according to a standard curve. Bacteria were washed twice with PBS and 2×10^8 cells/mL were incubated with soluble CEACAM-Fc or CEACAM-GFP proteins or CHO cell lysates for 1 h at 37 °C with head-over-head rotation. After incubation, bacteria were washed 5 times with PBS and either boiled in SDS sample buffer (62.5 mM Tris-HCl [pH 6.8], 2% w/v SDS, 10% glycerol, 50 mM DTT, and 0.01% w/v bromophenol blue) prior to SDS-PAGE and Western blotting or taken up in FACS buffer (PBS/0.5% BSA) for flow cytometry analysis.

Immunoprecipitation and Mass Spectrometry

Bacteria ($2x10^8$) in cold PBS containing protease and phosphatase inhibitors (Roche) were lysed by ultra-sonication on ice (10x, 20s). Cell debris was removed from the lysates by centrifugation at 15,000 rpm for 30 min at 4 °C, followed by pre-clearing with prewashed protein G-agarose (Roche Diagnostics). CEACAM1-Fc was added to the lysate (10 µg) and incubated for 1 h at 4 °C. Prewashed protein G-agarose (60 µL) were added to the antibody and lysate mixture and incubated 2 h at 4 °C. Beads were washed with PBS for five times to remove unspecifically bound proteins. Two-thirds of the beads were separated and used for mass spectrometry sample preparation. The supernatant was removed and the beads were resuspended twice in 50 µL 7M urea/ 2 M thiourea solved in 20 mM Hepes (pH 7.5) for denaturation of the proteins. Beads were pelleted by centrifugation and supernatants pooled and transferred to a new Eppendorf tube. Subsequently, proteins were reduced in 1 mM DTT for 45 min and alkylated at a final concentration of 5.5 mM iodacetamide for 30 min in the dark. The alkylation step was quenched by raising the DTT concentration to 5 mM for 30 min. All incubation steps were carried out at RT under vigorous shaking (Eppendorf shaker, 450 rpm). For digestion of the proteins 1 µL LysC (0.5 µg/µL) was added and the sample incubated for 4h at RT. To reduce the urea concentration the sample was diluted 1:4 with 50 mM triethylammonium bicarbonate and then incubated with 1.5 µL trypsin (0.5 µg/µL) at 37 °C over night. Trypsin was finally inactivated by acidification with formic acid. The supernatant was transferred to a new Eppendorf tube and pooled with the following wash fraction of the beads with 0.1% formic acid. The sample was adjusted to pH 3 with formic acid (100% v/v) and subjected to peptide desalting with a SepPak C18 column (50 mg, Waters). Briefly, the column was subsequently washed with 1 mL 100% acetonitrile and 500 µL 80% acetonitrile, 0.5% formic acid. The column was equilibrated with 1 mL 0.1% TFA, the sample was loaded and the column washed again with 1 mL 0.1% TFA. After an additional wash step with 500 µL 0.5% formic acid peptides were eluted twice with 250 µL 80% acetonitrile, 0.5% formic acid. The organic phase was then removed by vacuum centrifugation and peptides stored at -80 °C. Directly before measurement peptides were resolved in 20 µL 0.1% formic acid, sonified for 5 min (water bath) and the sample afterwards filtered with a prewashed and equilibrated filter (0.45 µm low protein binding filter, VWR International, LLC). Measurements were performed on an LC-MS system consisting of an Ultimate 3000 nano HPLC directly linked to an Orbitrap XL instrument (Thermo Scientific). Samples were loaded onto a trap column (2 µm, 100 A, 2 cm length) and separated on a 15 cm C18 column (2 µm, 100 A, Thermo Scientific) during a 150 min gradient ranging from 5

to 30% acetonitrile, 0.1% formic acid. Survey spectra were acquired in the orbitrap with a resolution of 60,000 at m/z 400. For protein identification up to five of the most intense ions of the full scan were sequentially isolated and fragmented by collision induced dissociation. The received data was analyzed with the Proteome Discoverer Software version 1.4 (Thermo Scientific) and searched against the *H. pylori* (strain G27) database (1501 proteins) in the SEQUEST algorithm. Protein N-terminal acetylation and oxidation of methionins were added as variable modifications, carbamidomethylation on cysteines as static modifications. Enzyme specificity was set to trypsin and mass tolerances of the precursor and fragment ions were set to 10 ppm and 0.8 Da, respectively. Only peptides that fulfilled X_{corr} values of 1.5, 2.0, 2.25 and 2.5 for charge states +1, +2, +3 and +4 respectively were considered for data analysis.

Cells, cell-bacteria co-culture and elongation phenotype quantitation assay

Gastric cancer cell lines MKN45[45], KatoIII (ATCC, HTB-103), MKN28 [46] and AGS (ATCC, CRL-1739) were obtained from ATCC and DSMZ, authenticated by utilizing Short Tandem Repeat (STR) profiling, cultured either sparse or to tight confluence in DMEM (GIBCO, Invitrogen, Carlsbad CA, USA) containing 2 mM L-glutamine (GIBCO, Invitrogen, CA, USA) supplemented with 10% FBS (GIBCO, Invitrogen, CA, USA) and 1% Penicillin/ Streptomycin (GIBCO, Invitrogen, CA, USA). All cell lines were maintained in an incubator at 37°C with 5% CO2 and 100% humidity, and were routinely mycoplasma-tested twice per year by DAPI stain and PCR. Plate-grown bacteria were suspended in DMEM and washed by centrifugation at 150 *g* for 5 min in a microcentrifuge. After resuspension in DMEM, the optical density at 600 nm was determined and bacteria were added to the overnight serum-deprived cells at different ratios of bacteria/cell (MOI) at 37°C to start the infection. After the indicated time, cells were washed twice with PBS and then lysed with 1% NP-40 in protease & phosphatase inhibitor PBS. HEK293 cells were chosen for CEACAM transfection studies because the cells were found to be negative for hu-CEACAM expression, and are easily transfectable. HEK cells were grown in 6-well plates containing RPMI 1640 medium (Invitrogen) supplemented with 25 mM HEPES buffer and 10% heat-inactivated FBS (Biochrom, Berlin, Germany) for 2 days to approximately 70% confluence. Cells were serum-deprived overnight and infected with *H. pylori* at MOI 50 for the indicated time points in each figure. After infection, the cells were harvested in ice-cold PBS containing 1 mM Na_3VO_4 (Sigma-Aldrich). Elongated AGS cells in each experiment were quantified in 5 different 0.25-mm^2 fields using an Olympus IX50 phase contrast microscope.

Transfection

A CHO cell line (ATCC) permanently expressing hu-CEACAM1-4L, mouse-CEACAM1-L and rat-CEACAM1-L were generated by stably transfecting cells with 4 µg pcDNA3.1-huCEACAM1-4L, pcDNA3.1-huCEACAM1-4S, pcDNA3.1-msCEACAM1-L, pcDNA3.1-ratCEACAM1-L plasmid (Singer), respectively, utilizing the lipofectamine 2000 procedure according to the manufacturer's protocol (Invitrogen). Stable transfected cells were selected in culture medium containing 1 mg/mL of Geniticinsulfat (G418, Biochrom, Berlin, Germany). The surface expression of CEACAM1 in individual clones growing in log phase was determined by flow cytometry (FACS calibur, BD). HEK293 cells were transfected with 4 µg of the HA-tagged CEACAM constructs or luciferase reporter constructs (Clontech, Germany) for 48 h with TurboFect reagent (Fermentas, Germany) according to the manufacturer's instructions.

Western blot

An equal volume of cell lysate was loaded on 8% SDS-PAGE gels and after electrophoresis, separated proteins were transferred to nitrocellulose membrane (Whatman/GE Healthcare, Freiburg, Germany). Membranes were blocked in 5% non-fat milk for 1 h at room temperature and incubated overnight with primary antibodies mAb 18/20 binding to CEACAM1,3,5, B3-17 and C5-1X (mono-specific for hu-CEACAM1, Singer), 4/3/17 (binding to CEACAM1,5, Genovac), and 5C8C4 (mono-specific for hu-CEACAM5, Singer), 1H7-4B (mono-specific for hu-CEACAM6, Singer), 6/40c (mono-specific for hu-CEACAM8, Singer), Be9.2 (α-rat-CEACAM1, kindly provided by Dr. W. Reutter, Charite, CBF, Germany), mAb 11-1H (α-rat-CEACAM1ΔN, Singer), phosphotyrosine antibody PY-99 (Santa Cruz, LaJolla, CA, USA), α-CagAphosphotyrosine antibody PY-972 [47], mouse monoclonal α-CagA antibody (Austral Biologicals, San Ramon, CA, USA), mouse monoclonal α-CEACAM1 (clone D14HD11Genovac/Aldevron, Freiburg,Germany) or goat α-GAPDH (Santa Cruz). After washing, membranes were incubated with the secondary antibody [HRP-conjugated α-mouse IgG (Promega)] and proteins were detected by ECL Western Blotting Detection reagents. The quantification was done by LabImage 1D software (INTAS).

Flow cytometry

The Fc-tagged CEACAMs (2.5 µg/mL) were incubated with *H. pylori* (OD_{600}=1) and subsequently with FITC-conjugated goat α-human IgG (Sigma-Aldrich). After washing with

FACS buffer, the samples were analyzed by gating on the bacteria (based on forward and sideward scatter) and measuring bacteria-associated fluorescence. In each case, 10,000 events per sample were obtained. Analysis was performed with the FACS CyAn (Beckman Coulter) and the data were evaluated with FlowJo software (Treestar). For the analysis of CEACAM mediated HopQ binding, indicated cell types (5×10^5 in 50 µL) were incubated with 20 µg/mL of *H. pylori* strain P12 derived,myc and 6x His-tagged recombinant HopQdiluted in 3% FCS/PBS for 1 h on ice. After three times washing with 3% FCS/PBS samples were labeled with 20 µg/mL of mouse α-c-mycmAb (clone 9E10, AbDSerotec) and subsequently with FITC conjugated goat α-mouse F(ab')2 (Dianova, Germany). In parallel, the presence of CEACAMs was controlled by staining cells utilizing the rabbit anti CEA pAb (A0115, Dianova) followed by FITC conjugated goat α-rabbit F(ab')2 (Dianova, Germany). Background fluorescence was determined using isotype-matched Ig mAb. The stained cell samples were examined in a FACScalibur flow cytometer (BD Biosciences, San Diego, CA) and the data were analyzed utilizing the CellQuest software. Dead cells, identified by PI staining, were excluded from the measurement.

Immunohistochemistry and Immunofluorescence

Following approval of the local ethics committee, paraffin-embedded human normal stomach, gastritis and cancer samples were randomly chosen from the tissue bank of the Institut für Pathologie, Klinikum Bayreuth Germany. Histological samples were excluded if tissue quality was poor. After antigen retrieval with 10 mM sodium citrate buffer pH 6 in pressure cooker, the sections were incubated with α-hu-CEACAM1, 5, 6 and α-rat-CEACAM1 antibodies (clone B3-17, 5C8C4, 1H7-4B and Be9.2, respectively). Sections were developed with SignalStain DAB (Cell Signaling) following manufacturer's instructions. Sections were counterstained with hematoxylin (Morphisto). The automated image acquisition was performed with Olympus Virtual Slide System VS120 (Olympus, Hamburg, Germany).

Visualization of the co-localization of HopQ and CEACAMs co-staining of normal and gastritis sections was performed utilizing HopQ-biotin followed by streptavidin-Cy3 and α-hu-CEACAM1, 3, 5, 6, 8 clone 6G5j followed by Alexa 488 coupled goat anti mouse antibody. The cell nuclei were stained with DAPI. DAPI and fluorescent proteins were analyzed with the Leica DMI4000B microscope.

Adherence assay

The adherence assay was performed according to Hytonen et al [48]. Briefly, human gastric epithelial cells (MKN45 and AGS) and CEACAM1-transfected CHO cells were grown in

antibiotic free DMEM (Gibco, Gaithersburg, MD) supplemented with 5% FCS and L-glutamine (2 mM, Sigma-Aldrich) on tissue culture 96 well plates (Bioscience) in 5% CO2 atmosphere for 2 days. To visualize *H. pylori* cells in adhesion assays, OD_{600}=1 of bacteria were fluorescence labeled with CFDA-SE (Molecular Probes) and washed with PBS. CFDA-SE was added at concentration of 10 μM for 30 min at 37˚C under constant rotation in the dark. Excess dye was removed by 3 times washing with PBS. Bacteria were resuspended in PBS until further use. Labelled bacteria were co-incubated (MOI 10) with the cells at 37°C with gentle agitation for 1 h. After washing with PBS (1 mL, ×3) to remove non-adherent bacteria, cells were fixed in paraformaldehyde (2%, 10 min). Bacterial binding was determined by measuring the percentage of cells that bound fluorescent-labeled bacteria using flow cytometry analysis.

IL-8 cytokine ELISA

AGS cell line was infected with *H. pylori* as described already and PBS-incubated control cells served as negative control. The culture supernatants were collected and stored at -20 °C until assayed. IL-8 concentration in the supernatant was determined by standard ELISA with commercially available assay kits (Becton Dickinson, Germany) according to described procedures.

HopQ-dependency of CagA virulence pathways

If not indicated otherwise, the AGS cell line (ATCC CRL-1730) was infected with the various *H. pylori* strains for 6 hours at a multiplicity of infection (MOI) of 50. The cells were then harvested in ice-cold PBS in the presence of 1 mM Na_3VO_4 (Sigma-Aldrich). In each experiment the number of elongated AGS cells was quantified in 10 different 0.25-mm^2 fields using a phase contrast microscope (Olympus IX50). CagA translocation was determined using the indicated antibodies detecting Tyr-phosphorylated CagA. All experiments were performed in triplicates. For inhibition experiments, cells were incubated with the indicated antibodies or peptides prior to infection.

Confocal microscopy

CHO cells were grown on chamber slides (Thermo Scientific), fixed in paraformaldehyde (4%, 10 min) and blocked with PBS/5% bovine serum albumin. CFDA-SE labelled bacteria (10 μM for 30 min at 37˚C under constant rotation in the dark) at MOI 5 were incubated with cells for 1 h at 37˚C under constant rotation. After 5X PBS washing, cell membranes were

stained with Deep Red (Life Technology) and cell nuclei with DAPI (Life Technology). Confocal images of cells were taken using a Leica SP5 confocal microscope.

Crystallization and structure determination of $HopQ^{AD}$

$HopQ^{AD}$was concentrated to 40 mg/mL and crystallized by sitting drop vapor diffusion at 20°C using 0.12 M alcohols (0.02 M 1,6-Hexanediol; 0.02 M 1-Butanol; 0.0 2M 1,2-Propanediol; 0.02 M 2-Propanol; 0.02 M 1,4-Butanediol; 0.02 M 1,3-Propanediol), 0.1 M Tris (base)/BICINE pH 8.5, 20% v/v PEG 500 MME; 10 % w/v PEG 20000 as a crystallization buffer. Crystals were loop-mounted and flash-cooled in liquid nitrogen. Data were collected at 100 K at beamline Proxima1 (SOLEIL, Gif-sur-Yvette, France) and were indexed, processed and scaled using the XDS package[49]. All crystals were in the $P2_1$ space group with approximate unit cell dimensions of a=57.7 Å, b=57.7 Å, c=285.7 Å and beta=90.1° and four copies of $HopQ_{442}$ per assymetric unit. Phases were obtained by molecular replacement using the BabA structure (PDB:5F7K)[21] and the program phaser[50,51]. The models were refined by iterative cycles of manual rebuilding in the graphics program COOT[52] and maximum likelihood refinement using Refmac5 [53]. Extended Data Table 2 summarizes the crystal parameters, data processing and structure refinement statistics.

Amino acid sequence alignment

The amino acid sequence alignment of the N-terminal domains of human, mouse and rat-CEACAM1 and human CEACAMs (1, 5, 6 and 8) was performed using CLC main Workbench (CLC bio).

Luciferase reporter assays

CHO-CEACAM1-L cells transfected with various luciferase reporter and control constructs (Clontech) were infected with *H. pylori* for 5 h and analyzed by luciferase assay using the Dual-Luciferase Reporter Assay System according to the manufactures instruction (Promega, USA). Briefly, cells were harvested by passive lysis, the protein concentration was measured with Precision Red (Cytoskeleton, USA) and the lysates were equalized by adding passive lysis buffer. The luciferase activity was measured by using a Plate Luminometer (MITHRAS LB940 from Berthold, Germany).

Animal experiments

Specific pathogen free, 120-150 g 4 weeks-old male Sprague Dawley rats, were obtained from Charles River Laboratories (Sulzfeld, Germany). Animals were randomly distributed into the different experimental groups by animal care takers not involved in the experiments, and

criteria for the exclusion of animals were pre-established. Investigator blinding was performed for all assessment of outcome and data, histology was performed by an independent investigator in a blinded manner. Animals were challenged twice intragastrically in groups of 8 with ~1 x 10^8 live *H. pylori* in 2 interval days. After 6 weeks infection, stomachs were removed and sectioned. One part was embedded in paraffin for histological analysis and another piece was weighted and homogenized to determine colony forming units (CFU)/mg stomach. Serial dilutions (1/10, 1/100 and 1/1000) were plated in WC dent plates. CFU were counted after 4 days.

The experiments were performed in the specific pathogen-free unit of Zentrum für Präklinische Forschung, Klinikum r. d. Isar der TU München, according to the allowance and guidelines of the ethical committee and state veterinary office (Regierung von Oberbayern, 55.2-1.54-2532-160-12).

Statistical Analysis

For in vitro experiments, normal distribution was determined by Shapiro–Wilk test. Normally distributed data were analyzed with two-tailed Student *t*-test or One-way ANOVA with post hoc Bonferroni test (comparing more than two groups) using Graph Pad Prism Software. Data are shown as mean ± s.e.m or S.D. for at least three independent experiments. P values <0.05 were considered significant. For animal studies, power calculation was performed based on previous animal experiments to achieve two sided significance of 0,05 while using lowest possible numbers to comply with the ethical guidelines for experimental animals. Mann-Whitney U test or ANOVA Kruskal-Wallis, Dunn's multiple comparison test were used to determine statistical significances.

References

1 Salama, N. R., Hartung, M. L. & Muller, A. Life in the human stomach: persistence strategies of the bacterial pathogen Helicobacter pylori. *Nature reviews. Microbiology* **11**, 385-399, doi:10.1038/nrmicro3016 (2013).

2 Atherton, J. C. & Blaser, M. J. Coadaptation of Helicobacter pylori and humans: ancient history, modern implications. *The Journal of clinical investigation* **119**, 2475-2487, doi:10.1172/JCI38605 (2009).

3 Montecucco, C. & Rappuoli, R. Living dangerously: how Helicobacter pylori survives in the human stomach. *Nature reviews. Molecular cell biology* **2**, 457-466, doi:10.1038/35073084 (2001).

4 Linden, S., Mahdavi, J., Hedenbro, J., Boren, T. & Carlstedt, I. Effects of pH on Helicobacter pylori binding to human gastric mucins: identification of binding to non-MUC5AC mucins. *The Biochemical journal* **384**, 263-270, doi:10.1042/BJ20040402 (2004).

5 Ilver, D. *et al.* Helicobacter pylori adhesin binding fucosylated histo-blood group antigens revealed by retagging. *Science* **279**, 373-377 (1998).

6 Mahdavi, J. *et al.* Helicobacter pylori SabA adhesin in persistent infection and chronic inflammation. *Science* **297**, 573-578, doi:10.1126/science.1069076 (2002).

7 Solnick, J. V., Hansen, L. M., Salama, N. R., Boonjakuakul, J. K. & Syvanen, M. Modification of Helicobacter pylori outer membrane protein expression during experimental infection of rhesus macaques. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 2106-2111, doi:10.1073/pnas.0308573100 (2004).

8 Hammarstrom, S. The carcinoembryonic antigen (CEA) family: structures, suggested functions and expression in normal and malignant tissues. *Seminars in cancer biology* **9**, 67-81, doi:10.1006/scbi.1998.0119 (1999).

9 Obrink, B. On the role of CEACAM1 in cancer. *Lung cancer* **60**, 309-312, doi:10.1016/j.lungcan.2008.03.020 (2008).

10 Gray-Owen, S. D. & Blumberg, R. S. CEACAM1: contact-dependent control of immunity. *Nature reviews. Immunology* **6**, 433-446, doi:10.1038/nri1864 (2006).

11 Voges, M., Bachmann, V., Kammerer, R., Gophna, U. & Hauck, C. R. CEACAM1 recognition by bacterial pathogens is species-specific. *BMC microbiology* **10**, 117, doi:10.1186/1471-2180-10-117 (2010).

12 Heneghan, M. A. *et al.* Effect of host Lewis and ABO blood group antigen expression on Helicobacter pylori colonisation density and the consequent inflammatory response. *FEMS immunology and medical microbiology* **20**, 257-266 (1998).

13 Virji, M., Watt, S. M., Barker, S., Makepeace, K. & Doyonnas, R. The N-domain of the human CD66a adhesion molecule is a target for Opa proteins of Neisseria meningitidis and Neisseria gonorrhoeae. *Molecular microbiology* **22**, 929-939 (1996).

14 Hill, D. J. & Virji, M. A novel cell-binding mechanism of Moraxella catarrhalis ubiquitous surface protein UspA: specific targeting of the N-domain of carcinoembryonic antigen-related cell adhesion molecules by UspA1. *Molecular microbiology* **48**, 117-129 (2003).

15 Kuespert, K., Roth, A. & Hauck, C. R. Neisseria meningitidis has two independent modes of recognizing its human receptor CEACAM1. *PloS one* **6**, e14609, doi:10.1371/journal.pone.0014609 (2011).

16 Peek, R. M. Helicobacter pylori infection and disease: from humans to animal models. *Disease models & mechanisms* **1**, 50-55, doi:10.1242/dmm.000364 (2008).

17 Icatlo, F. C., Goshima, H., Kimura, N. & Kodama, Y. Acid-dependent adherence of Helicobacter pylori urease to diverse polysaccharides. *Gastroenterology* **119**, 358-367 (2000).

18 Cao, P. & Cover, T. L. Two different families of hopQ alleles in Helicobacter pylori. *Journal of clinical microbiology* **40**, 4504-4511 (2002).

19 Ohno, T. *et al.* Relationship between Helicobacter pylori hopQ genotype and clinical outcome in Asian and Western populations. *J Gastroenterol Hepatol* **24**, 462-468, doi:10.1111/j.1440-1746.2008.05762.x (2009).

20 Alm, R. A. *et al.* Comparative genomics of Helicobacter pylori: analysis of the outer membrane protein families. *Infection and immunity* **68**, 4155-4168 (2000).

21 Moonens, K. *et al.* Structural Insights into Polymorphic ABO Glycan Binding by Helicobacter pylori. *Cell host & microbe* **19**, 55-66, doi:10.1016/j.chom.2015.12.004 (2016).

22 Rossez, Y. *et al.* The lacdiNAc-specific adhesin LabA mediates adhesion of Helicobacter pylori to human gastric mucosa. *The Journal of infectious diseases* **210**, 1286-1295, doi:10.1093/infdis/jiu239 (2014).

23 Singer, B. B. *et al.* Deregulation of the CEACAM expression pattern causes undifferentiated cell growth in human lung adenocarcinoma cells. *PloS one* **5**, e8747, doi:10.1371/journal.pone.0008747 (2010).

24 Muenzner, P., Bachmann, V., Zimmermann, W., Hentschel, J. & Hauck, C. R. Human-restricted bacterial pathogens block shedding of epithelial cells by stimulating integrin activation. *Science* **329**, 1197-1201, doi:10.1126/science.1190892 (2010).

25 Slevogt, H. *et al.* CEACAM1 inhibits Toll-like receptor 2-triggered antibacterial responses of human pulmonary epithelial cells. *Nature immunology* **9**, 1270-1278, doi:10.1038/ni.1661 (2008).

26 Belogolova, E. *et al.* Helicobacter pylori outer membrane protein HopQ identified as a novel T4SS-associated virulence factor. *Cell Microbiol* **15**, 1896-1912, doi:10.1111/cmi.12158 (2013).

27 Mahler, M. *et al.* Experimental Helicobacter pylori infection induces antral-predominant, chronic active gastritis in hispid cotton rats (Sigmodon hispidus). *Helicobacter* **10**, 332-344, doi:10.1111/j.1523-5378.2005.00320.x (2005).

28 Chang, Y. J. *et al.* Mechanisms for Helicobacter pylori CagA-induced cyclin D1 expression that affect cell cycle. *Cell Microbiol* **8**, 1740-1752, doi:10.1111/j.1462-5822.2006.00743.x (2006).

29 Muenzner, P., Naumann, M., Meyer, T. F. & Gray-Owen, S. D. Pathogenic Neisseria trigger expression of their carcinoembryonic antigen-related cellular adhesion molecule 1 (CEACAM1; previously CD66a) receptor on primary endothelial cells by activating the immediate early response transcription factor, nuclear factor-kappaB. *The Journal of biological chemistry* **276**, 24331-24340, doi:10.1074/jbc.M006883200 (2001).

30 Olbermann, P. *et al.* A global overview of the genetic and functional diversity in the Helicobacter pylori cag pathogenicity island. *PLoS genetics* **6**, e1001069, doi:10.1371/journal.pgen.1001069 (2010).

31 Suerbaum, S. & Josenhans, C. Helicobacter pylori evolution and phenotypic diversification in a changing host. *Nature reviews. Microbiology* **5**, 441-452, doi:10.1038/nrmicro1658 (2007).

32 Baltrus, D. A. *et al.* The complete genome sequence of Helicobacter pylori strain G27. *Journal of bacteriology* **191**, 447-448, doi:10.1128/JB.01416-08 (2009).

33 Arnold, I. C. *et al.* Tolerance rather than immunity protects from Helicobacter pylori-induced gastric preneoplasia. *Gastroenterology* **140**, 199-209, doi:10.1053/j.gastro.2010.06.047 (2011).

34 Lee, A. *et al.* A standardized mouse model of Helicobacter pylori infection: introducing the Sydney strain. *Gastroenterology* **112**, 1386-1397 (1997).

35 Lundin, A. *et al.* The NudA protein in the gastric pathogen Helicobacter pylori is an ubiquitous and constitutively expressed dinucleoside polyphosphate hydrolase. *J Biol Chem* **278**, 12574-12578, doi:10.1074/jbc.M212542200 (2003).

36 Atherton, J. C. *et al.* Mosaicism in vacuolating cytotoxin alleles of Helicobacter pylori. Association of specific vacA types with cytotoxin production and peptic ulceration. *The Journal of biological chemistry* **270**, 17771-17777 (1995).

37 Cover, T. L., Dooley, C. P. & Blaser, M. J. Characterization of and human serologic response to proteins in Helicobacter pylori broth culture supernatants with vacuolizing cytotoxin activity. *Infect Immun* **58**, 603-610 (1990).

38 Backert, S., Muller, E. C., Jungblut, P. R. & Meyer, T. F. Tyrosine phosphorylation patterns and size modification of the Helicobacter pylori CagA protein after translocation into gastric epithelial cells. *Proteomics* **1**, 608-617, doi:10.1002/1615-9861(200104)1:4<608::AID-PROT608>3.0.CO;2-G (2001).

39 Vermoote, M. *et al.* Genome sequence of Helicobacter suis supports its role in gastric pathology. *Vet Res* **42**, 51, doi:10.1186/1297-9716-42-51 (2011).

40 Haesebrouck, F. *et al.* Non-Helicobacter pylori Helicobacter species in the human gastric mucosa: a proposal to introduce the terms H. heilmannii sensu lato and sensu stricto. *Helicobacter* **16**, 339-340, doi:10.1111/j.1523-5378.2011.00849.x (2011).

41 Schott, T., Kondadi, P. K., Hanninen, M. L. & Rossi, M. Comparative genomics of Helicobacter pylori and the human-derived Helicobacter bizzozeronii CIII-1 strain reveal the molecular basis of the zoonotic nature of non-pylori gastric Helicobacter infections in humans. *BMC Genomics* **12**, 534, doi:10.1186/1471-2164-12-534 (2011).

42 Tegtmeyer, N. *et al.* Characterisation of worldwide Helicobacter pylori strains reveals genetic conservation and essentiality of serine protease HtrA. *Molecular microbiology* **99**, 925-944, doi:10.1111/mmi.13276 (2016).

43 Singer, B. B. *et al.* Soluble CEACAM8 interacts with CEACAM1 inhibiting TLR2-triggered immune responses. *PLoS One* **9**, e94106, doi:10.1371/journal.pone.0094106 (2014).

44 Studier, F. W. Protein production by auto-induction in high density shaking cultures. *Protein expression and purification* **41**, 207-234 (2005).

45 Hojo, H. & Onishi, Y. [Case suspected to be atypical diffuse myeloma]. *Nihon rinsho. Japanese journal of clinical medicine* **35**, 2659-2662 (1977).

46 Romano, M., Razandi, M., Sekhon, S., Krause, W. J. & Ivey, K. J. Human cell line for study of damage to gastric epithelial cells in vitro. *The Journal of laboratory and clinical medicine* **111**, 430-440 (1988).
47 Mueller, D. *et al.* c-Src and c-Abl kinases control hierarchic phosphorylation and function of the CagA effector protein in Western and East Asian Helicobacter pylori strains. *The Journal of clinical investigation* **122**, 1553-1566, doi:10.1172/JCI61143 (2012).
48 Hytonen, J., Haataja, S. & Finne, J. Use of flow cytometry for the adhesion analysis of Streptococcus pyogenes mutant strains to epithelial cells: investigation of the possible role of surface pullulanase and cysteine protease, and the transcriptional regulator Rgg. *BMC Microbiol* **6**, 18, doi:10.1186/1471-2180-6-18 (2006).
49 Krauth-Siegel, R. L. *et al.* Crystallization and preliminary crystallographic analysis of trypanothione reductase from Trypanosoma cruzi, the causative agent of Chagas' disease. *FEBS letters* **317**, 105-108 (1993).
50 Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta crystallographica. Section D, Biological crystallography* **67**, 235-242, doi:10.1107/S0907444910045749 (2011).
51 McCoy, A. J. *et al.* Phaser crystallographic software. *Journal of applied crystallography* **40**, 658-674, doi:10.1107/S0021889807021206 (2007).
52 Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta crystallographica. Section D, Biological crystallography* **66**, 486-501, doi:10.1107/S0907444910007493 (2010).
53 Murshudov, G. N. *et al.* REFMAC5 for the refinement of macromolecular crystal structures. *Acta crystallographica. Section D, Biological crystallography* **67**, 355-367, doi:10.1107/S0907444911001314 (2011).

Acknowledgments

We thank Jeannette Koch, Judith Lind, Birgit Maranca-Hüwel and Bärbel Gobs-Hevelke for their excellent technical support; Carolin Konrad, Johannes Fischer for support with rat experiments and Marie Roskrow for fruitful discussion and revision. KM and HR acknowledge use of the Soleil synchrotron, Gif-sur-Yvette, France under proposal 20131370 and support by VIB and the Flanders Science Foundation (FWO) through the Odysseus program, a postdoctoral fellowship and Hercules funds UABR/09/005. This work was supported by the German Centre for Infection Research, partner site Munich, to MG, by the BMBF 01EO1002 to E.K., the Mercator Research Center Ruhr An2012-0070 to BBS, the German Science Foundation CRC-796 (B10) and CRC-1181 (A04) to SB, the Collaborative Research Center/Transregio 124, Project A5 to HS.

Author Contribution

A.J., T.K., K.M., N.T., B.K., N.B., A.S. and B.B.S performed the experiments, B.B.S, R.H., V.K., E.K., H.S. and C.R.H. provided reagents and tools, A.J., B.B.S, H.R., D.B., R.M.-L., S.B. and M.G. conceived the experiments, analyzed the data and wrote the manuscript. All authors read and approved the final manuscript.

Author information

Reprints and permissions information is available at www.nature.com/reprints. M.G., B.K. and T.K. are employees and Shareholders of Imevax GmbH. M.G., A.J., B.S., S.B. and T.K. are named as inventors on a patent application regarding HopQ. The other authors declare no conflict of interest. Correspondence and requests for materials should be addressed to markus.gerhard@tum.de.

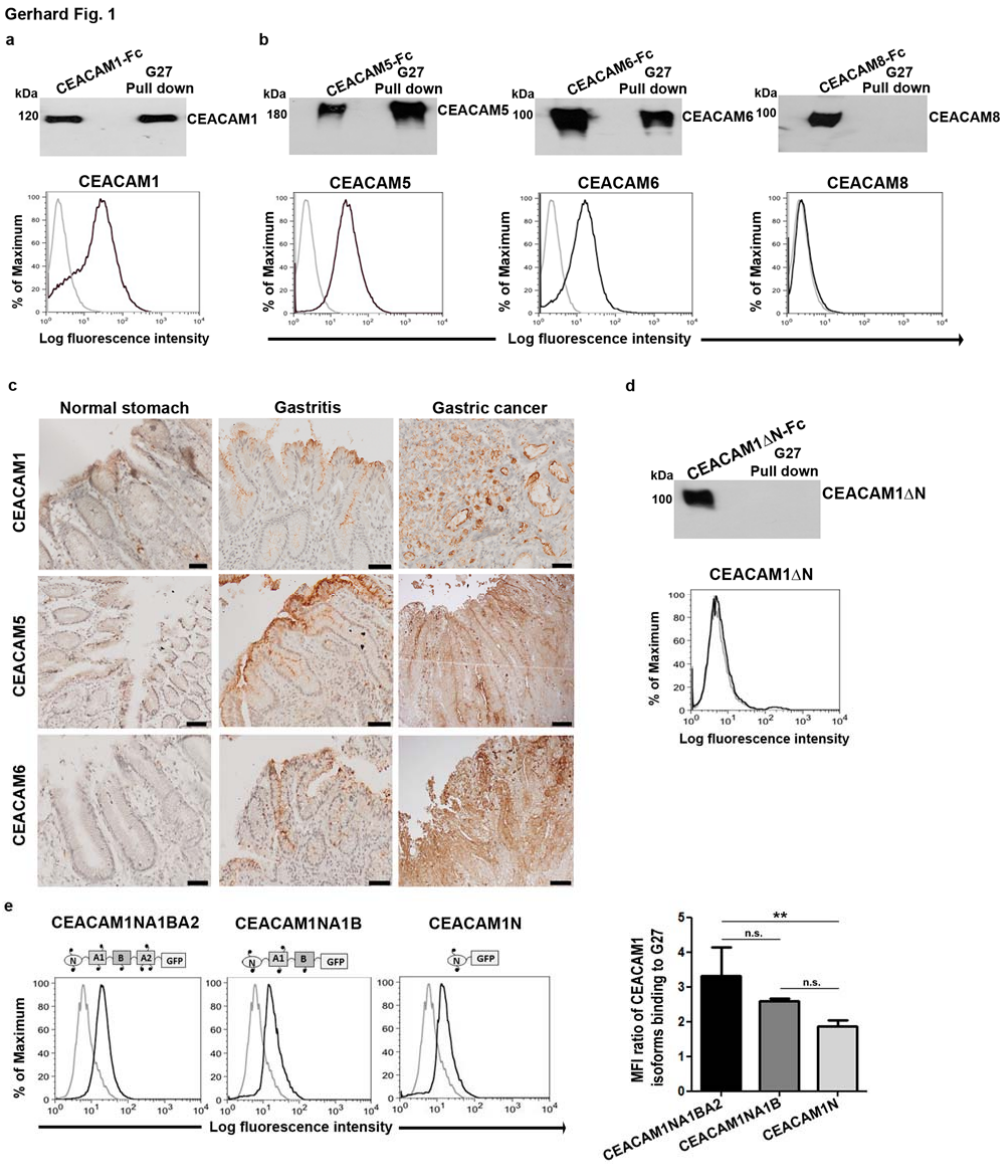
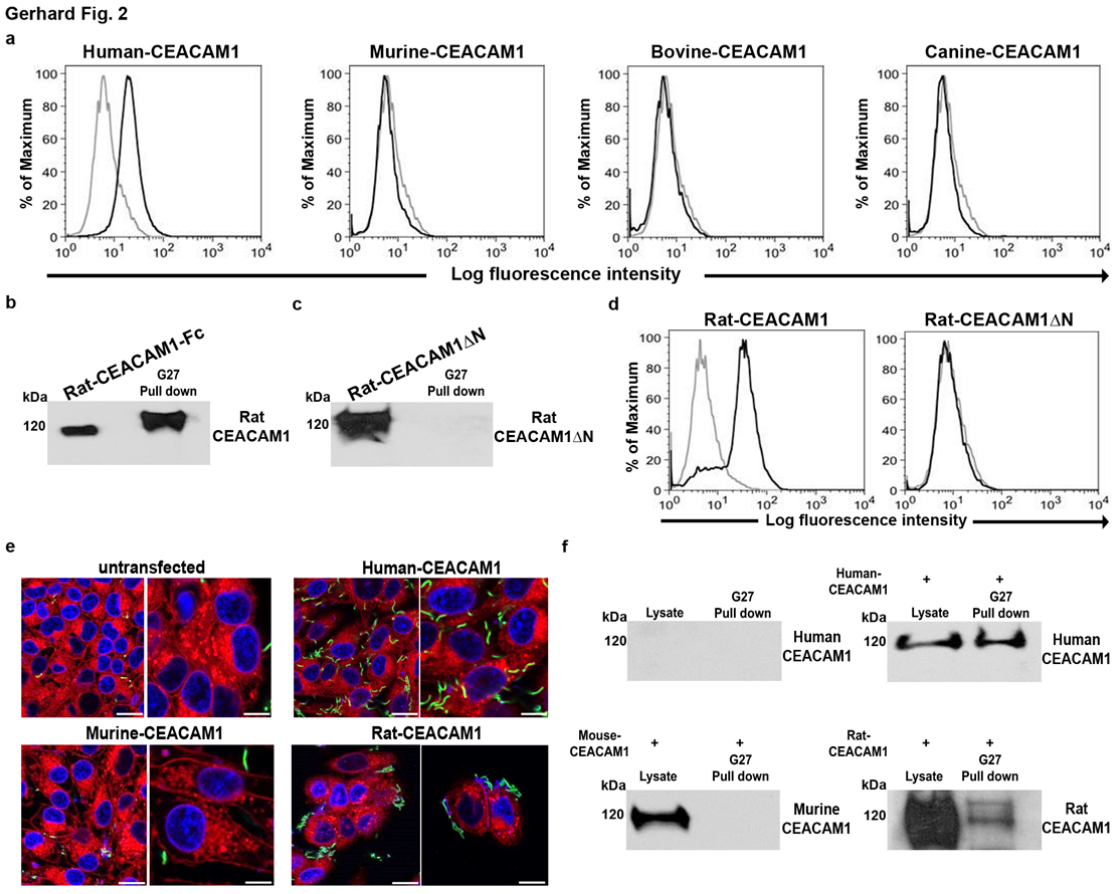


Figure 1 *H. pylori* employs the N-terminal domain of hu-CEACAM1 and binds CEACAM5 and CEACAM6 but not CEACAM8. *H. pylori* G27 strain binding to human CEACAM1-Fc (a) and human CEACAM5-Fc, CEACAM6-Fc or CEACAM8-Fc (b) was analyzed by pull down experiments followed by western blot analysis and flow cytometry (n=3). (c) CEACAM1, CEACAM5 and CEACAM6 expression detected by immunohistochemistry in human normal stomach, gastritis and gastric cancer samples**.** Scale bars, 50 µm. (d) Binding of *H. pylori* to human CEACAM1ΔN-Fc (lacking the complete N-domain) detected by western blot after pull down or by flow cytometry. One representative experiment of 4 is shown. (e) *H. pylori* binding to CEACAM variants analyzed by flow cytometry. Mean Fluorescence Intensity (MFI) ratios (mean, S.D.) are shown (n=4). One-way ANOVA, *P* value= 0.009, n. s.: not significant.



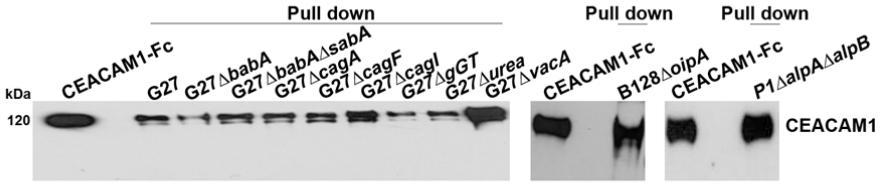
878

879

880 **Figure 2 *H. pylori* binding to CEACAM1 orthologues.** (a) *H. pylori* G27 strain binding to
881 human, murine, bovine and canine CEACAM1 determined by flow cytometry. (b) and (c) *H.*
882 *pylori* (G27) binding to rat-CEACAM1-Fc (b) and rat-CEACAM1ΔN-Fc (c) detected by
883 western blot after bacterial pull down. (d) Binding of G27 *H. pylori* strain to rat-CEACAM1
884 and rat-CEACAM1ΔN detected by flow cytometry. (e) Representative confocal images of *H.*
885 *pylori* binding to human, rat and mouse CEACAM1-expressing CHO cells. Untransfected
886 CHO served as control. Scale bars: left panels, 25 µm, right panels, 10 µm. (f) *H. pylori* G27
887 pull down of whole cell lysates of untransfected, human-, mouse- and rat CEACAM1-
888 transfected CHO cells. CEACAM1 was detected using species-specific CEACAM1
889 antibodies, as indicated. Representative experiments are shown (n=3).

Gerhard Fig. 3

a

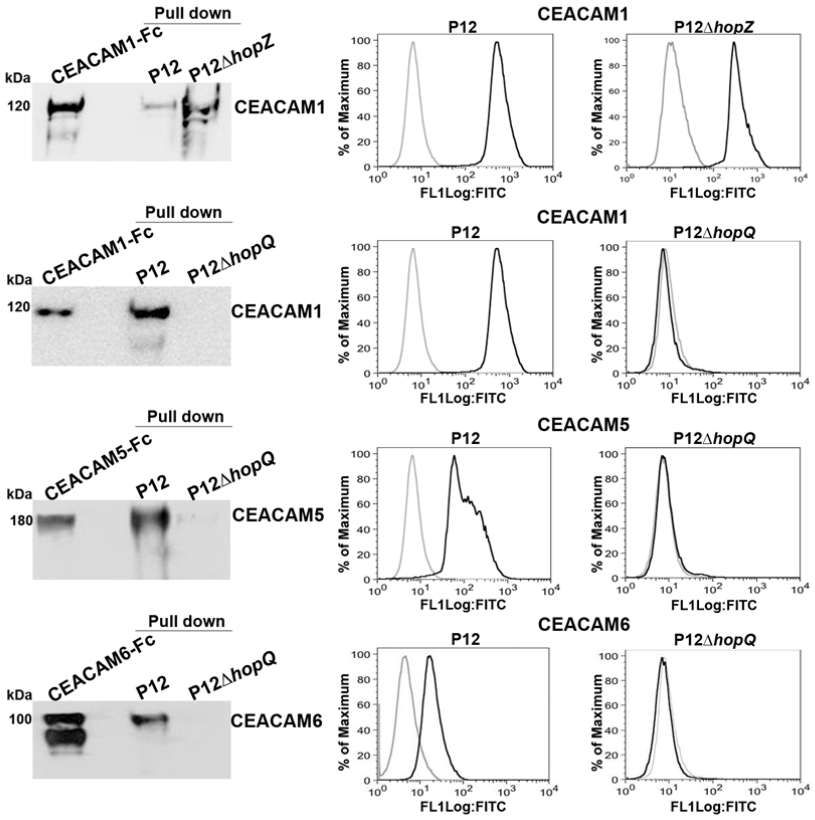


b

Mass spectrometry analysis

Accession	Description	Score	Unique Peptides	PSMs	Area	AAs	MW [kDa]
B5Z8H1	Outer membrane protein hopQ	34.29	5	9	4.714E8	641	69.9
B5Z6D0	Outer membrane protein hopZ	31.74	4	8	7.565E8	666	72.4
B5Z783	Outer membrane protein sabB	23.37	5	7	8.180E8	638	70.7
B5Z7R8	Outer membrane protein hopC	14.40	3	3	6.037E8	517	56.2
B5ZA84	Outer membrane protein babA	8.0	1	2	4.238E8	744	80.8
B5Z786	Outer membrane protein sabA	6.66	2	2	2.580E8	651	72.1

c



890

891 **Figure 3 *H. pylori* binds to CEACAM1 via HopQ.** (a) Human CEACAM1 detected by
892 western blot after pull down of various *H. pylori* G27 knockout strains incubated with human
893 CEACAM1-Fc. (b) Candidate outer membrane proteins of *H. pylori* strain G27 binding to
894 human CEACAM1-Fc (for complete MS table see Suppl. Table 1). (c) *H. pylori* strains P12,
895 P12Δ*hopQ* and P12Δ*hopZ* binding to hu-CEACAM1-, CEACAM5- and CEACAM6-Fc
896 detected by western blot and FACS analysis after pull down. Representative experiments are
897 shown (n=3).

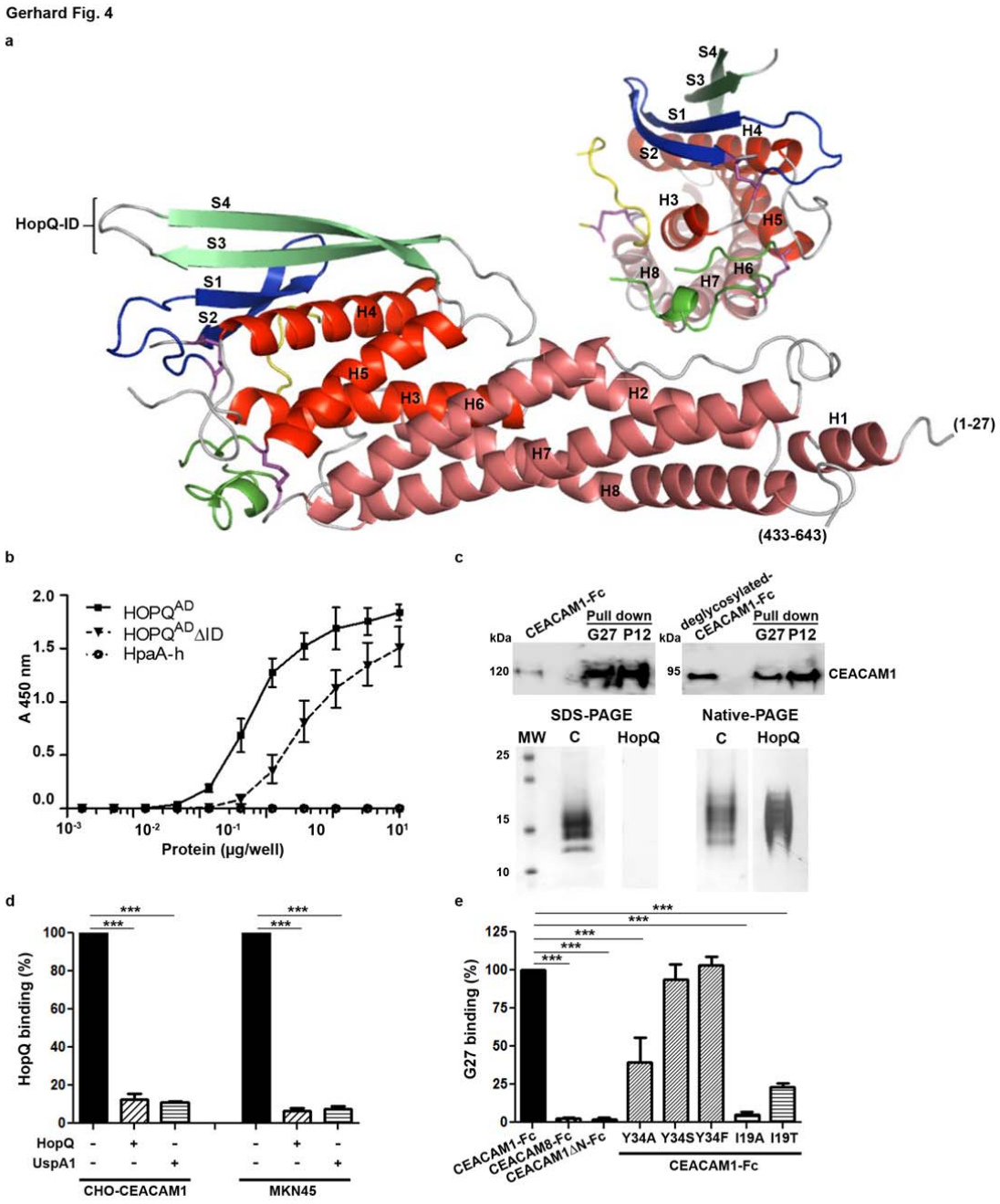


Figure 4. X-ray structure and binding properties of the HopQ adhesin domain. (a) Ribbon representation of the HopQ^{AD} showing the 3+4-helix bundle topology (colored red and brick, respectively). Three Cys pairs (Cys102-Cys131, Cys237-Cys269 and Cys361-Cys384) conserved in most Hop family members pinch off extended loops are colored blue, yellow and green. HopQ-ID; green, β-hairpin insertion. (b) ELISA titers of HopQ^{AD} or mutant HopQ^{AD} lacking the HopQ-ID (HopQ^{AD}ΔID) binding to increasing concentrations of C1-N domain (C1ND) (*n*=4, mean, S.D.). (c) Upper panel, pull down experiments of *H. pylori* strains incubated with de-glycosylated human CEACAM1-Fc. Lower panel, SDS and native

907 page of C1ND stained with Coomassie-blue ("C") or with $HopQ^{AD}$ in a far western blot
908 ("HopQ") experiment. (d) HopQ binding (%) to CEACAM1 in CHO and MKN45 cells after
909 pre-incubation with recombinant HopQ or UspA1, respectively. Mean, S.D. of three
910 independent experiments are shown. (e) *H. pylori* G27 binding (%) to CEACAM1,
911 CEACAM1ΔN and different CEACAM1 variants. CEACAM8 was used as negative control.
912 Mean, S.D. of three independent experiments are shown. One-way ANOVA with Bonferroni's
913 correction for multiple comparisons. ***$P \leq 0.001$.
914
915
916
917
918

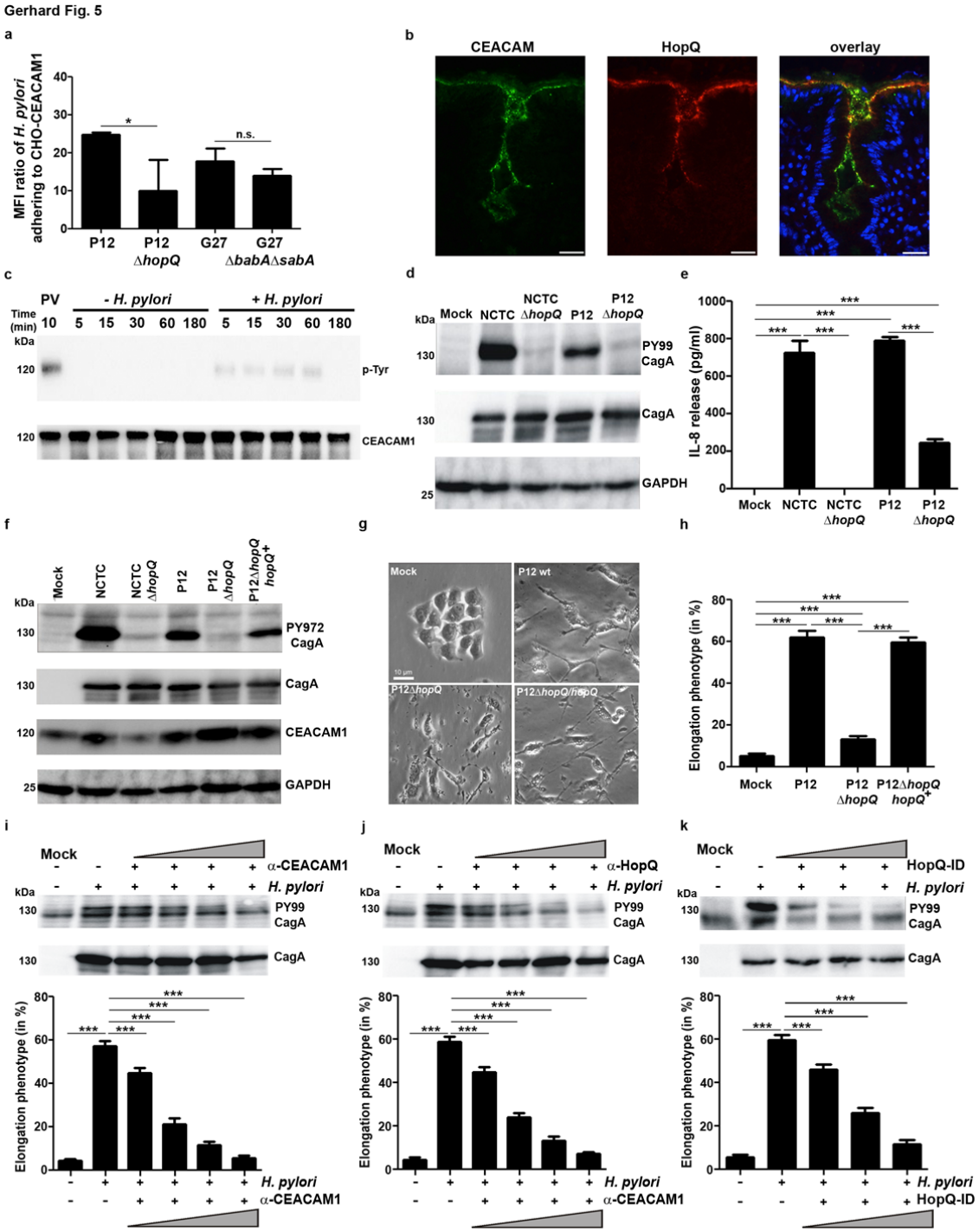


Figure 5 Deletion of *hopQ* in *H. pylori* leads to reduced bacterial cell adhesion and abrogates CagA delivery, IL-8 release and cell elongation. (a) *H. pylori* binding to CHO-hu-CEACAM1-L cells detected by flow cytometry analysis (n=3). Means ± S.D. are shown. Two-tailed *t*-test, * $P \leq 0.03$. (b) Immunofluorescence detection of apical CEACAM expression (green) and HopQ binding (red) in the gastric epithelium from human gastritis biopsies. Scale bar 25 µm. (c) CEACAM1 Tyr-phosphorylation and total CEACAM1 levels in

uninfected and *H. pylori*-infected CHO-CEACAM1-L cells. Pervanadate (PV) treatment served as positive control. (d) CagA phosphorylation detected in lysates of AGS cells after infection with *H. pylori* P12, NCTC11637 and corresponding isogenic *hopQ* mutants (e) Secreted IL-8 by AGS cells after infection with the indicated *H. pylori* strains (mean, S.D. of three independent experiments are shown). One-way ANOVA with Bonferroni's correction for multiple comparisons. ***P≤0.001. (f) CagA phosphorylation and CEACM1 levels in HA-tagged HEK293-hu-CEACAM1 transfectants infected with indicated *H. pylori* strains. (g) Representative phase contrast micrographs of AGS cells infected for 6 h with P12, P12Δ*hopQ* or P12Δ*hopQhopQ*$^{+}$ re-expressing wt *hopQ* gene. (h) Quantification of elongation phenotype induced in AGS cells after infection with the indicated *H. pylori* strains. Data (mean, S.D.) of three independent experiments are shown. One-way ANOVA with Bonferroni's correction for multiple comparisons. ***P≤0.001. (i) CagA phosphorylation and quantification of the elongation phenotype (five different 0.25-mm^2 fields) after *H. pylori* P12 infection of AGS cells pre-treated with 2, 5, 10 or 20 µg of α-CEACAM Ab (lanes 3-6). Data (mean, S.D.) of three independent experiments are shown. One-way ANOVA with Bonferroni's correction for multiple comparisons. ***P≤0.001. (j) CagA phosphorylation and quantification of the elongation phenotype after infection of AGS with wild type *H. pylori* pre-treated with 2, 5, 10 or 20 µg of α-HopQ (lanes 3-6) Data (mean, S.D.) of three independent experiments are shown. One-way ANOVA with Bonferroni's correction for multiple comparisons. ***P≤0.001. (k) CagA phosphorylation in *H. pylori*-infected AGS cells pre-incubated with a HopQ-derived peptide (1 µM, 2.5 µM and 5 µM) corresponding to the HopQ-ID (aa 189-220). Cell elongation (mean, S.D.) from 3 independent experiments is shown. One-way ANOVA with Bonferroni's correction for multiple comparisons. ***P≤0.001.

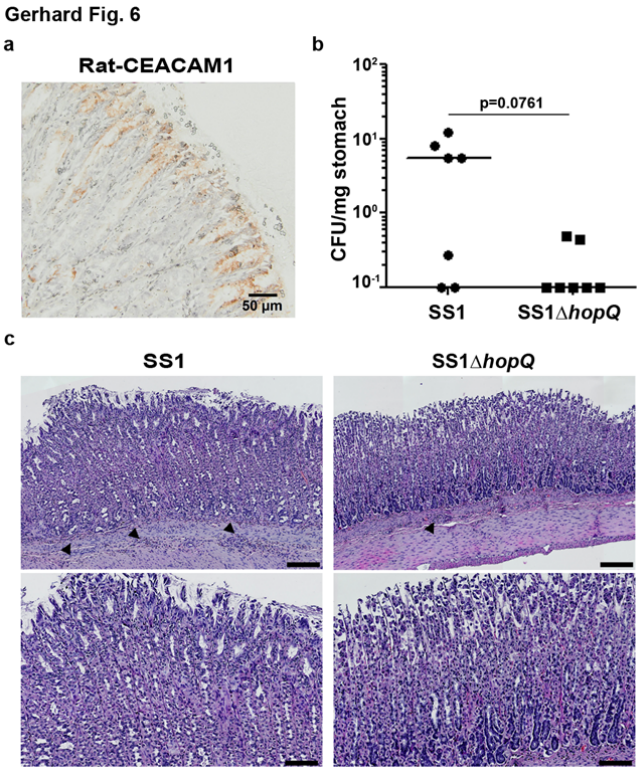


Figure 6 *H. pylori* colonization of rat stomach depends on HopQ. (a) CEACAM1 expression in rat stomach. (b) *H. pylori* colony forming units (CFU) per mg stomach of male Sprague dawley rats after 6 weeks infection. Horizontal bars indicate medians. Mann-Whitney U test. (c) Hematoxylin/eosin staining of infected rat stomachs. Representative images of same stomach regions are shown. Scale bar 100µm (upper panels) and 200µm (lower panels). Arrows denote inflammatory cells.